REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author HINKS J A.

COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

Studies of Retroviral Proteases.

John Andrew Hinks

A thesis submitted for the degree of Doctor of Philosophy

Department of Biochemistry and Molecular Biology,

University College London,

Gower Street,

London,

WC1E 6BT.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This work is primarily concerned with the expression, purification, and characterisation of aspartic proteases from three retroviruses of the lentivirus subgroup, specifically the Human Immunodeficiency Viruses types 1 and 2, and the Simian Immunodeficiency Virus isolated from the African Green Monkey (HIV-1 PR, HIV-2 PR, SIV_{AGM} PR respectively). These viruses cause immunodeficiency syndromes within their respective hosts, and understanding their molecular biology would facilitate development of Acquired Immunodeficiency Syndrome (AIDS) treatments in man. The proteases are essential to viral maturation and infection; and are of great interest with respect to the development of new antivirals.

This thesis describes attempts to develop improved methods for the over-expression and purification of these cytotoxic proteins for use in structural and biochemical studies. The development of a system for expression and purification of active, crystallisable HIV-1 PR is described, followed by a preliminary analysis of two compounds intended to act as irreversible "suicide inhibitors" of HIV-1 PR. The expression, purification, crystallisation, and preliminary crystallographic data for the native HIV-2 PR using the same expression system are also reported.

Mutagenesis of the HIV-2 protease is described, whereby the conserved active site aspartic acids of the native homodimer were changed to histidine and cysteine, with the intention of modifying the enzyme's mechanism, whilst maintaining its native substrate specificity and overall structure.

Finally the production of an insoluble, non-toxic, histidine tagged fusion of mutant *E.coli* Uracil DNA Glycosylase (UDG) and the SIV_{AGM} PR is reported. This was intended to produce high levels of a non-toxic fusion protein, allow one-step affinity purification, and provide native soluble protease following autocatalytic cleavage from the fusion protein. The effects of protease toxicity and codon usage on yields are discussed in light of the results presented.

Acknowledgements

I would like to thank my supervisors, Professor Laurence Pearl, and Dr. Eric DeSouza for their support and training. I would like to thank Dr Chris Prodromu, Dr. Stuart Wilson, and Dr. Renos Savva for their helpful practical advice and for many enlightening discussions. I would also like to thank Dr. Chris Richardson, Dr. Chris Gray, and Dr. Jane Wibley for their technical help during the preparation of this thesis. I am especially grateful to Dr. Laura Pickles for completing the onerous task of proof reading the entire manuscript, and for invaluable encouragement.

I pay compliments to all my friends and colleagues past and present for their cheer and comradeship over the years, and thank my family for their moral sustenance. Finally I would like to say a very special thank you to Shradha, for her unstinting support in every respect throughout.

This work was supported by a BBSRC CASE award and carried out at the Biochemistry and Molecular Biology Department of University College London, Gower street, London, and Wellcome Research Laboratories, Beckenham, Kent during the period from January 1992 to August 1996.

Table of Contents

Abbreviations

The table below defines the abbreviations used in this text. Note that abbreviations for chemical elements were as designated by international convention, whilst SI (System International) symbols were used for abbreviation of physical units. The conventional single and three letter codes for amino acids have both been used. Many commonly used scientific abbreviations (such as DNA, RNA) are not listed here since it was assumed that these would not require explaination. In any case each abbreviation is follwed by an explaination at point of first use within the text. Note that genes are typed in lower case, whilst proteins are typed in upper case (eg. tat and TAT).

Table of abbreviations.

AI	Angiotensin I
AII	Angiotensin II
AGM	African Green Monkey
AIDS	Acquired Immuno Deficiency Syndrome
Amp	Ampicillin
APS	Ammonium PerSulphate
ARC	AIDS Related Complex
ARV-2	AIDS-associated retrovirus-2
ATP	Adenosine TriPhosphate
AUC	Analytical Ultra Centrifugation
AZT	Azidothymidine
B-TrCP	[beta]-transducin repeat-containing protein
Bg	Background
CA	Capsid protein
CCP4	Suite of Crystallographic programs
CD	Circular Dichroism
CD4+	Receptor found on surface of T cells
CDC	Centres for Disease Control
Chlor	Chloramphenicol
Chpt	Chapter
CIP	Calf Intestinal Phosphatase
CIAP	Calf Intestinal Alkaline Phosphatase
CPM	Counts Per Minute
CTL	Cytotoxic T-Lymphocyte
CTD	C-Terminal Domain
CypA	Cyclophilin A
DAN	DiazoAcetyl Norleucine methyl ester
ddw	Double Distilled Water
DIQ	Decahydro-IsoQuinoline
DMS	DiMethylSilane
DMSO	DiMethylSulphOxide
dNTPs	deoxy-Nucleotide-TriPhosphates
DTNB	5,5'-dithio-bis(2-nitrobenzoic acid) (Ellman's reagent)
DTT	DiThioThreitol
EDTA	EthyleneDiamineTetra Acetate
EF-TU-GTP complex	Elongation Factor – Temperture Unstable: Guanosine 5'-Triphosphate complex
EIAV	EquineInfectious Aneamia Virus
EMBL	European Molecular Biology Laboratory
EPNP	1,2-Epoxy-3-(p-nitrophenoxy) propane
ER	Endoplasmic Reticulum
FDA	Federal Drugs Administration
FPLC	Fast Protein Liquid Chromatography ™
fig (s)	figure (s)
Frag	Fragment
GCG	Genetics Computer Group (Bioinformatics software package)
gp	glycol protein (eg. gp41 or gp120)
GST	Glutathione S-Transferase
HIC	Hydrophobic Interaction Chromatograph

His-tag Histidine tag

HIV Human Immuno Deficiency Virus

HLA Human Leukocyte Antigen

HTLV-III Human T-cell Lymphotropic Virus III

IEC Ion Exchange Chromatography

ILL Insitut Laue-Langevin

IMAC Immobilised Metal Afinity Chromatography

IN Integrase

IPTG Isopropyl β-d-thiogalactoside

IR Infra Red (spectroscopy)

IUBMB International Union of Biochemistry and Molecular Biology

K_{cat} Rate constant

K_M Michaelis-Menton constant

lac I Gene encoding the lac repressor

lac Iq lac I "quantity" mutant

LAV Lymphadenopathy-Associated Virus

LB Luria Broth

LBHB Low-Barrier Hydrogen Bond

LTR Long Terminal Repeat

MA Matrix protein

MBP Maltose Binding Protein

MES 2-Morpholino-Ethane Sulfonic acid

2ME 2-MercaptoEthanol

MHCI Major Histocompatibility Complex I

MHR Major Homology Region

MIP Mono Isopropyl-Phosphate

MK MegaKaryocytes

MS Mass Spectroscopy

MWCO Molecular Weight Cut-Off

Mwt Molecular Weight

NC NucleoCapsid

NEB New England Biolabs

NEF Negative Factor

NF-kB Nuclear Factor-kB

Ni-NTA Nickel Nitrilotriacetic acid

NMR Nuclear Magnetic Resonance (spectroscopy)

NNRTI Non-NRTI

NRTI Nucleoside / Nucleotide Reverse Transcriptase Inhibitor

OD_{280} Optical Density at λ=280 nm

OD_{600} Optical Density at λ= 600 nm

ORF Open Reading Frame

PAGE Poly-Acrylamide Gel Electrophoresis

PBMC Peripheral Blood Mononuclear Cells

PBS Phosphate Buffered Saline

PCR Polymerase Chain Reaction

PDB Protein Data Base

PEG Poly Ethylene Glycol

Pfu Pyrococcus furiosus

PHA Phytohemaglutanin

PMA Phorbol-12-myristate-13-acetate

PMSF Phenyl methyl sulfonyl fluoride

PR Protease

QSB Quick Screen Buffer

RCF Relative Centrifugal Force

REV Regulator of Expression of Viral proteins

RIL *E.coli* strain encoding tRNA for rare codons of Arg, Iso and Leu.

RPM Revolutions Per Minute

RT Reverse Transcriptase

SDS Sodium Dodecyl Sulphate

SEC Size Exclusion Chromatography

SI Suicide Inhibitor

SIV Simian Immunodeficiency Virus

SOB Buffered *E.coli* growth medium

SP1 Specificity protein 1 (a zinc finger transcription factor).

SPA Scintillation Proximity Assay

SRS Synchrotron Radiation Source

SSHB Short Strong Hydrogen Bond

Sta Statine

SU Surface envelope protein

T4 Lymphocytes aka T-helper or CD4+ cells

T7 Promoter from Bacteriophage T7

TAR Trans-Activating Response element

TAT Transactivator protein

Taq Thermus aquaticus

TB Transformation Buffer

TBE Tris Borate EDTA

TBP TATA-box binding protein

TE Tris EDTA

TEMED NNN'N'Tetramethyl-1,2-ethanediamine

TM Transmembrane protein

T_m Melting temperature

Trc Hybrid of Trp and Lac promoters

Tris Tris Hydroxymethylaminoethane

Trp Promoter from *E.coli* Tryptophan operon

TWEEN (Tween 20) Registered Trademark

U Uracil

UDG Uracil DNA Glycosylase

UF Ultra Filtration

UNAIDS United Nations AIDS program

UV Ultra Violet

Vent trademark of thermophilic DNA polymerase

vis visible

VPR Viral Protein R

VPU Viral Protein U

VPX Viral Protein X

VIF Viral Infectivity Factor

V_{max} Maximum velocity

WHO World Health Organisation

List of tables

Appendix

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix

Chapter One: General Introduction.

The Human Immunodeficiency Virus (HIV) (Gallo, Wong-Staal *et al.* 1988) is the etiological agent of the Acquired Immunodeficiency Syndrome (AIDS) (Weiss, Hollander *et al.* 1985; Gallo and Montagnier 1988), a disease that has evolved into a world-wide public health problem (Barre-sinoussi, Chermann *et al.* 1983; Gallo, Salahuddin *et al.* 1984; Pinching and Weiss 1986). The need to develop effective therapeutic agents against the virus is well recognised and a brief summary of the retroviral life cycle is provided in this chapter to allow an appreciation of the potential targets. As this project began all licenced drugs for the treatment of AIDS were based on nucleoside analogues, targeting the enzyme Reverse Transcriptase (RT) (Mitsuya, Yarchoan *et al.* 1990). The drug of choice was Zidovudine (AZT), first approved by the US Food and Drug Administration (FDA) for treatment of AIDS in 1987. There are now thirteen NRTI (Nuceloside / Nucleotide Reverse Transcriptase Inhibitor) formulations, and three NNRTI (Non-NRTI) compounds approved by the FDA (Federal Drugs Administration) and commercially available. However, toxicities associated with AZT (Richman, Fischl *et al.* 1987) and the demonstration of AZT resistant strains of virus (Larder, Darby *et al.* 1989; Larder and Kemp 1989; Boucher, O'Sullivan *et al.* 1992) - problems also encountered with later anti-RT drugs - led researchers to focus on other potential antiviral targets. The most promising of these alternatives appeared to be the viral aspartic protease (PR) (Wlodawer, Miller *et al.* 1989) encoded by the pol open reading frame (Fauci 1988) of HIV, which is essential for the proteolytic processing of both the gag and the gag-pol polyprotein precursors during replication (Kramer, Schaber *et al.* 1986; Debouck, Gorniak *et al.* 1987; Graves, Lim *et al.* 1988; Le Grice, Mills *et al.* 1988; Mous, Heimer *et al.* 1988; Seelmeier, Schmidt *et al.* 1988). Inactivation of this protease, by biochemical or genetic means, results in the production of immature, non-infectious virions (Kohl, Emini *et al.* 1988; Peng, Ho *et al.* 1989). This project aimed to contribute towards an increased understanding of the structures, mechanisms and inhibition of the retroviral proteases derived from HIV-1, HIV-2 and SIV_{AGM}. There are now five PR inhibitors commercially available, all developed by means of rational drug design strategies (reviewed in Ren and Lien 1998; Ren and Lien 2001). Such strategies were only possible once high resolution three dimensional structural solutions had been obtained for these proteases and their inhibitor complexes. A major aim of this project was to

obtain such data for the HIV-2 PR apoenzyme, a task that proved intractable. Structural studies demand high levels of the target protein but the retroviral proteases are notoriously difficult to express from recombinant protein expression systems. Early structural successes relied variously on synthetic protein (Schneider, J. and S. B. Kent 1988; Wlodawer, Miller *et al.* 1989) or expression on a very large scale (approximately 3000 litre fermentations) (Danley, Geoghegan, *et al.* 1989; Lapatto, Blundell *et al.* 1989; Darke, Leu, *et al.* 1989; McKeever, Navia *et al.* 1989). This project intended to address the problem by development of strategies for SIV_{AGM} PR overexpression in bacteria. It resulted in the construction of a new bacterial expression system able to produce the retroviral protease as a non-toxic and inactive fusion protein, which can be converted to a soluble and fully functional protease once extracted from the bacterial cell. Development of this system has shed further light on the probable causes of low protease expression in bacteria and other expression systems.

A collaboration with an organic chemist produced compounds, which successfully inhibited HIV-1 PR. These compounds had been designed using a three dimensional computer model of the protease based on available structural data, and were intended either to disrupt dimerisation or to permanently block the protease active site. Their efficacy and mode of action was investigated using HIV-1 PR expressed in bacteria and purified to homogeneity for this purpose.

Mutants of HIV-2 PR were successfully made which allowed the creation of a heterodimeric PR having an active site intended to simulate that of a cysteine protease (Papain).

Aims of this study.

- To establish in house methods of large scale expression and purification for each of the proteases, or optimise them if already in existence (HIV-1), such that sufficient protein could be obtained to allow pursuit of the further aims.
- To attempt crystallographic studies of HIV-2 protease with a view to solving its structure, thus facilitating the rational design of protease inhibitors.
- To attempt the mutagenesis of the HIV-2 protease active site such that its mechanism of action might be altered (from that of an aspartate protease to that of papain), perhaps furthering our understanding of proteolytic mechanisms in general.
- To attempt crystallographic studies of SIV_{AGM}, protease with a view to solving its structure.
- To determine the efficacy of novel protease inhibitors designed to bind covalently to the active site of HIV-1 protease.

As will be seen, an overexpression and purification system for both HIV-1 PR (chapter 3) and HIV-2 PR (Chapter 4) was established. Sufficient HIV-1 PR was available to allow a preliminary study of two rationally designed protease inhibitors (Chapter 3). The HIV-2 PR apoenzyme was crystallised and some data collected (Chapter 4). However, no structural solution was possible from the available data and the collection of an entire data set was prevented by the rapid degradation of crystals on their exposure to X-rays.

Mutagenesis of the HIV-2 PR active site was carried out (Chapter 5) both to study the effect loss of activity might have on protease expression, and to create a heterodimeric mutant protease which was predicted to show proteolytic activity by virtue of a pseudo-papain active site.

Following the publication of an HIV-2 PR structure (Tong, Pav *et al.* 1993; Chen, Li *et al.* 1994), attention was turned to SIV_{AGM}. An examination of conventional methods for SIV_{AGM} protease expression and purification was made (Chapter 6), which resulted in the development of a novel expression system for this protease

(Chapter 7). The purified SIV_{AGM} was characterised and efforts to crystallise it were made. Structures of SIV protease were subsequently published by other groups (Wilderspin and Sugrue 1994; Rose, Craik *et al.* 1996) though not from the African Green Monkey serotype.

As a consequence of work on both the HIV-2 PR mutants and SIV_{AGM} expression system a revised view of the factors effecting bacterial expression of retroviral proteases was proposed and investigated, and suggestions made for future work (Chapter 8).

AIDS is a global health problem.

AIDS (Acquired Immune Deficiency Syndrome) was first discovered as a new disease in the USA, by the Centres for Disease Control (CDC) in 1981 (Sellers 1982). In 1983/84 French (Barre-sinoussi, Chermann *et al.* 1983) and American (Gallo, Salahuddin *et al.* 1984; Wong-Staal, Hahn *et al.* 1984) investigators (lead by Montagnier and Gallo respectively) identified the retrovirus now known as HIV (Human Immunodeficiency Virus) and confirmed it as the etiological agent of AIDS (Goedert and Gallo 1985). HIV was initially referred to by one of a number of names including Lymphadenopathy-Associated Virus (LAV), Human T-cell Lymphotropic Virus III (HTLV-III), and AIDS-associated Retrovirus-2 (ARV-2) until a consensus was reached (Biberfeld, Brown *et al.* 1987; Gallo, Wong-Staal *et al.* 1988). The virus was first successfully cultured in 1987 (Article: Nature, News 1987), there were two serotypes and each was shown to induce AIDS (Essex and Kanki 1988) (Marlink, Ricard *et al.* 1988).

In this chapter I will outline the lifecycle, structure and genomic organisation of HIV. In doing so aspects of its molecular biology will be introduced, their roles in viral infectivity discussed, and the possibility of targeting AIDS therapies against them considered. My intentions are first to provide a general overview of HIV research, and second to give an insight into the area of protease research in particular. It is not my intention to deal with AIDS *per se*, its distribution, or any related diseases - other than as an initial justification for the study of HIV, and in order to illustrate the ways in which HIV affects its host. In addition the importance of $CD4^+$ T-lymphocytes (T_4 cells, or T-helper cells) to the normal functioning of the immune system will be assumed to be prior knowledge, and is amply covered in the text of Riott's Essential Immunology (Roitt and Delves 2001). It is also worth noting that for the purposes of this chapter no distinction will be made between HIV-1 and HIV-2, and the term 'HIV' will be used to include both serotypes, unless one serotype is specifically stipulated.

On its recognition in 1981, AIDS was seen to be typified by outbreaks of Kaposi's Sarcoma, and *Pneumocystis carinii* in previously healthy homosexual men. By mid-1988 100 000 cases of AIDS had been reported to the World Health Organisation (WHO) from 136 countries, and it was estimated that 500 000 to 1 million people were infected by HIV. Dr. J. Mann, head of the WHO special program on AIDS at

that time, considered that the HIV infected population worldwide might actually number 5 to 10 million, while the director of the CDC Dr. James Curran, claimed that one new case of AIDS was being diagnosed every 14 minutes. Since that time the global epidemic has continued to grow, and in the years between 1997 and 2003 the number of new infections has been estimated at approximately 5 million per year worldwide. At the end of 2004 it is expected that approximately 28 million AIDS deaths will have been reported worldwide since the outbreak of the epeidemic[1].

Initially the majority of reported infections occurred in the USA, yet AIDS is global. Thousands of cases have been reported in Europe, but the greatest incidence is now in Africa (particularly sub-Saharan Africa) where only a fraction of cases are reported to the WHO. Figures for 2003 from the WHO and UNAIDS suggest that there are between 34 and 46 million people currently living with the infection (although some estimates are as high as 60 million), of whom 60-70% (25-28 million) are in sub-Saharan Africa, and about 10-20% (4-8 million) are in southern Asia. Throughout the world, more epicentres are emerging where there had previously been very few cases of HIV, such as in China, Russia or the Ukraine. With the cost of treatment currently estimated at around $10 000 (£6250) per patient per annum the majority of infections in Asia and Africa go untreated. This combined with poor education and a lack of appropriate preventative measures has ensured that levels of infection continue to rise. A total of 14 million children are estimated to have been orphaned through AIDS, and in 16 countries south of the Sahara, 10% or more of 15 to 49 year olds have become infected with HIV. The highest incidence is in Botswana, where one adult in three carries the virus, and two-thirds of all boys now aged 15 will die prematurely of AIDS. In South Africa, where a fifth, and in Zimbabwe, where a quarter of all adults have become infected, AIDS will kill half of all 15-year olds.

The effect in areas worst hit has been social and economic catastrophe. In Zambia, 1300 teachers died in the first ten months of 1998, equivalent to two-thirds of all the

[1] Figures herein are based on the most recent available data from UNAIDS (UNAIDS AIDS Epidemic Update 2003). UNAIDS/WHO publish updated country estimates biannually. The next update will be released in advance of the International AIDS Conference, to be held this year in Bangkok (11-16 July 2004). Until then, the country estimates available are from end 2001 (published in the "Report on the Global HIV/AIDS Epidemic 2002"). The most recent regional and global estimates were published in the "AIDS Epidemic Update, December 2003".

country's newly qualified teachers. In Zimbabwe AIDS infections and related deaths have resulted in a 61% reduction in rural maize production. Across the South African mining industry an average of 25% of mine workers are infected. Recently Zambia's largest cement company reported a 15-fold increase in the number of workers taking leave in order to attend funerals. These and other anecdotes are reminiscent of a plague-infested Europe in the middle ages. In the absence of preventative measures or affordable treament the situation will presumably worsen, and have a knock on effect on the societies worst hit as labour shortages plus a break down in social and governmental administration take their toll.

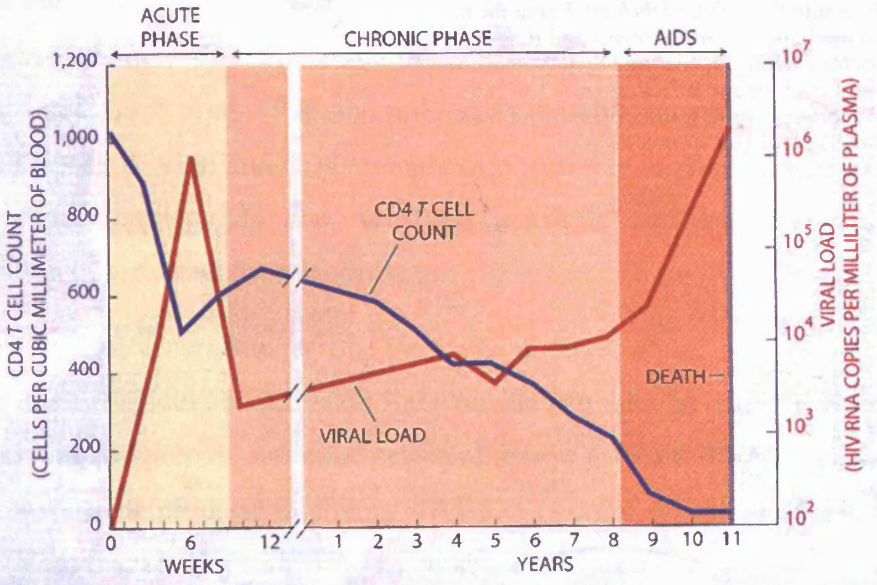
HIV infection is spread by sexual contact, by infected blood or blood products, and perinatally by mother to child. It can damage organs directly, but its crippling effect on the immune system also allows the development of opportunistic infections - which account for up to 90% of deaths from AIDS (AIDS being the final stage of HIV infection – see fig. 1, 'Progression through stages of HIV infection'). HIV impairs immune function through its elimination of T-lymphocytes and macrophages (representing the cell-mediated arm of the immune system). Cytotoxic T lymphocytes (CTLs) provide a highly effective anti-viral defense mechanism, and can detect virus-infected cells (by recognition of viral peptides presented at the cell surface within the HLA class I binding cleft) and kill them. These cells are effective against HIV infection in the short to medium term (Ogg, Jin *et al.* 1998; Schmitz, Kuroda *et al.* 1999), but ultimately cannot control infection. These cells are known to carry the $CD4^+$ receptor (discussed later, see 'Fusion and entry') and as illustrated in figure 1, declining $CD4^+$ cell numbers can be correlated with progression of the syndrome (although some immune dysfunction may be present before $CD4^+$ cell numbers are seen to decline, so that there may be some other mechanism involved).

A healthy person has approximately 1000 $CD4^+$ cells per ml of blood. This number drops by 40 to 80 cells per ml each year in an HIV infected individual. Once the level has fallen below 200 to 400 cells per ml, many of the opportunistic infections associated with 'AIDS Related Complex' (ARC) appear. The major killers are *Pneumocystis carinii*, Crytococcal meningitis, and toxoplasmosis. These account for 50 to 70% of deaths. The risk of *P.carinii* increases greatly once the $CD4^+$ cell count drops below 200 per ml, and risk of the others increases when the count falls below 100 per ml. Evasion of the CTL response by HIV has been the subject of investigation

(Phillips, Rowland-Jones *et al.* 1991; Evans, O'Connor *et al.* 1999) and the positive selection of mutations within the viral sequences encoding the CTL-recognised epitopes may be responsible. These mutations significantly reduce the ability of CTLs to kill infected cells.

Figure 1. Progression through stages of HIV infection.

This figure is taken from Scientific American (July 1998 p66), but the original source was Fauci, Pantaleo *et al.* 1996.



SOURCE: Anthony Fauci et al. in Annals of Internal Medicine, *Vol. 124; 1996*

NATURAL COURSE OF HIV INFECTION in a typical untreated patient begins with a sharp rise of virus in the blood (*orange line*) and a consequent drop in CD4 *T* cells (*blue line*), the immune cells most damaged by HIV. The immune system soon recovers somewhat, however, and keeps HIV levels fairly steady for several years. Eventually, though, the virus gains the upper hand. AIDS is diagnosed when the CD4 *T* cell level drops below 200 cells per cubic millimeter of blood or when opportunistic infections (reflecting failed immunity) arise, whichever happens first.

66 SCIENTIFIC AMERICAN July 1998

The point at which $CD4^+$ count dips below 400 cells per ml is marked by the onset of ARC, and is coincident with an equivalent titre of viral particles. Shortly after this point the viral titre increases unabated as the $CD4^+$ count plummets towards the 200 cells per ml mark, where AIDS is officially diagnosed. The chronic phase (Fig. 1) was first thought to represent a period of viral latency during, since infected persons did not necessarily present any obvious symptoms of disease. However, the measurement of viral load and $CD4^+$ cell count indicated that this might not be the case. As the work described in this thesis began (1992) the mechanism behind the observed decay

in $CD4^+$ count was not clear, and it has since been the subject of extensive research. This area of study has impinged on, and benefited from, developments in drug design - especially advances in protease inhibitor design between 1992 and 1997. Consequently it will be considered here to provide a backdrop to the work of this thesis, and to demonstrate the contribution made to the understanding of viral / host interaction, and to AIDS therapy in general, by the development of protease inhibitors.

By 1993 it had been demonstrated that increased HIV-1 load correlated with a reduction in $CD4^+$ lymphocyte levels and consequent disease progression (Weiss 1993; Wain-Hobson 1993). The importance of viral replication to the disease state was clear (Wei, Gosh *et al.* 1995 and references therein), however, little was known about the kinetics of viral and $CD4^+$ lymphocyte turnover *in vivo* because sufficiently potent antiviral compounds had not been available and because methods for quantification of virus had been inadequate.

Two studies (Ho, Neumann, *et al.* 1995; Wei, Gosh *et al.* 1995) were able to overcome these difficulties and shed light on the problem by exploiting new more potent antiviral compounds and more sensitive assays for viral RNA. The compounds in question were the protease inhibitors ABT-538 (Kempf, Marsh *et al.* 1995) and L-735,524 (Vacca, Dorsey *et al.* 1994; Dorsey, Levin *et al.* 1994), and the reverse transcriptase inhibitor nevirapine (Merluzzi, Hargrave *et al.* 1990).

The clinical trial of the protease inhibitor ABT-538 (Markowitz, Mo *et al.* 1995) provided an opportunity for Ho *et al.* to determine what steady state (if any) existed between viral and $CD4^+$ lymphocyte turnovers *in vivo* by disrupting that state. Prior to and during the treatment of trial patients $CD4^+$ lymphocyte counts and HIV-1 levels were monitored, and the findings reported in Nature (Ho, Neumann *et al.* 1995). The results suggested that approximately half the plasma virions turned-over every two days – inferring that HIV-1 replication *in vivo* must be high. Furthermore, on average the entire population of $CD4^+$ lymphocytes was seen to be turning over every fifteen days in patients prior to treatment and thus before disruption of any steady state. Since on treatment $CD4^+$ counts were seen to rise they concluded that the depletion of CD4+ cells during AIDS could be attributed to $CD4^+$ cell destruction by HIV-1. They estimated that in an infected patient approximately 10^9 $CD4^+$ cells are destroyed and

replenished daily, and commented that this was close to the number of HIV-1 RNA expressing lymphocytes calculated to be present in the body by *in situ* PCR and hybridisation experiments (Embretson, Zupancic *et al.* 1993; Haase1994).

Wei *et al.* (Wei, Gosh *et al.* 1995) also used clinical trials on ABT-538 to study the kinetics of viral production and clearance, but additionally employed the results of trials on L-735,524, and nevirapine as well (see review of HIV protease inhibitors, this chapter). Besides measuring viral load changes by PCR analysis of plasma borne viral RNA, they also quantified changes in the viral genotype and phenotype by employing automated DNA sequencing and an *in situ* assay of reverse transcriptase function. Their results concurred with those of Ho *et al.*, and they concluded that the majority of circulating plasma virus was derived from "continuous rounds of *de novo* infection, replication and cell turnover, and not from cells that produce virus chronically or are latently infected and become activated" (Wei, Gosh *et al.* 1995). They also calculated that in order for steady-state levels of the virus to be maintained the total viral population must be replaced daily (i.e. something like 1.1×10^8 virions per day), whilst $CD4^+$ cells were turned over at a rate of about 2×10^9 cells per day. Remarkably their clinical trials also showed that wild type virus was almost completely replaced by drug resistant virus in the plasma within 14 days.

In summing up Ho *et al.* (Ho, Neumann *et al.* 1995) famously used a "sink" analogy to emphasise that the $CD4^+$ depletion seen at later stages of infection was due to virus production finally outstripping the rate at which $CD4^+$ cells could be replenished. The period of supposed 'latency' between initial infection and the onset of AIDS was in fact shown to be a period of extreme, yet finely balanced, activity on behalf of both virus and host. A dynamic 'steady-state' in which viral production and clearance, and $CD4^+$ cell infection and replenishment were balanced had lately been anticipated (Nowak, Anderson *et al.* 1991; Coffin 1992; Wain-Hobson 1993) but the coincidently published works of Ho, Wei and respective co-workers confirmed the departure from earliest thinking, which had supposed that the virus lay 'dormant' (hence lentivirus or slow virus) for many years until 'awakened' by some extracellular stimulus to the host cells, or more recently that HIV specific CD8 T-lymphocytes were responsible for maintaining a low viral load and thus allowing a long period of asymptomatic infection (Buseyne and Riviere 1993; Pantaleo, Demarest *et al.* 1994). Importantly Wei *et al.* also pointed the way to the future therapeutic strategy of combinatorial

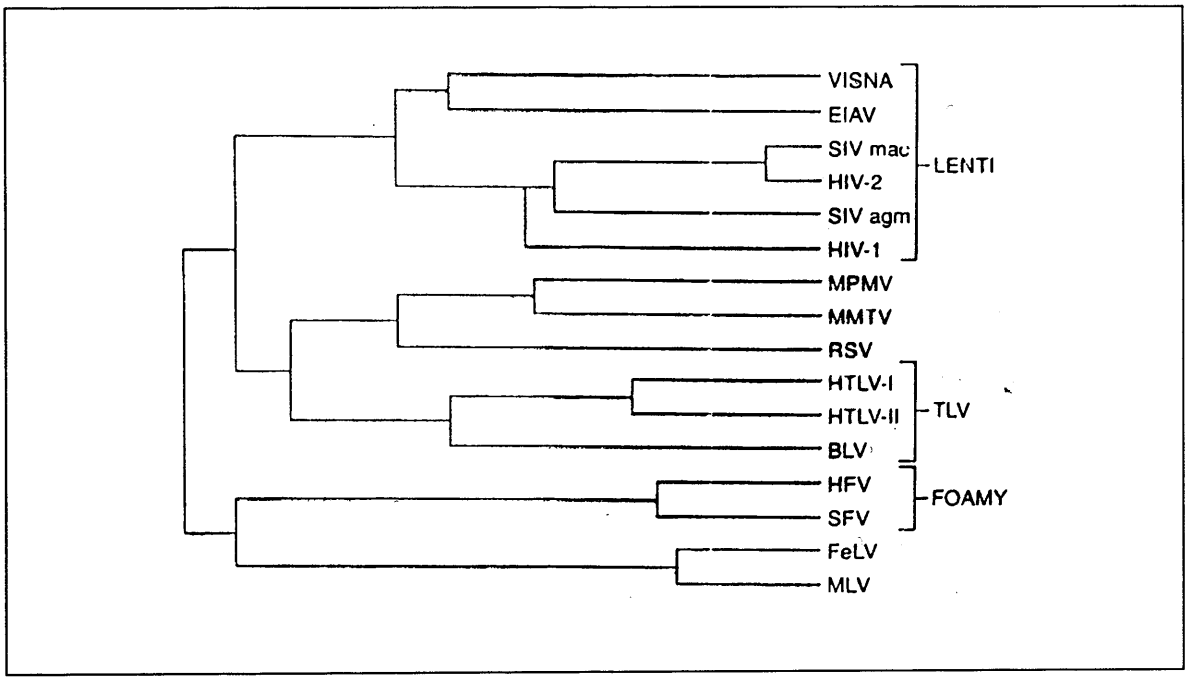
therapy, since they demonstrated that the virus could very quickly develop resistance to compounds delivered individually.

The papers provoked a flurry of comments and replies (Wain-Hobson 1995; Nowak *et al.* 1995; Mosier 1995; Bukrinsky, Manogue *et al.* 1995; Phillips, Sabin *et al.* 1995; Ascher, Sheppard *et al.* 1995; Michie 1995; Buianouckas 1995; Weber and Galpin 1995; Cohen 1995; James 1995; Coffin 1995; Levy, Ramachandran *et al.* 1996), not all in agreement (Duesberg and Bialy 1995). Duesberg and Bialy considered that the mathematical models were flawed, and that in any case the results merely confirmed that HIV was a passenger virus rather than the causal agent of AIDS. Despite this further study reiterated and refined the initial findings (Perelson, Neumann *et al.* 1996). Overall these results had revitalised belief that a patient's immune system might recover if HIV reproduction could be stopped long enough, even though the virus could not be completely eradicated. They had also validated the use of small, rapid clinical trials and the employment of PCR based measurements of plasma borne viral RNA to measure viral titres *in vivo*, and they had emphasised the importance of combinatorial therapy.

The retrovirus family

HIV and SIV are members of the family of viruses known as the Retroviridae, thus they are retroviruses. The defining characteristic of this family is a genome comprising RNA which is reverse transcribed to DNA by viral reverse transcriptase. Retroviridae are currently divided into seven genera, including the Lentiviruses. This genus incorporates retroviruses with complex genomes and cone shaped capsid core particles, enveloped by a lipid bi-layer derived from their host's cell membrane. The Lentiviruses are subdivided into five species - the Visna Virus, the Equine Infectious Anemia Virus (EIAV), the Feline Immunodeficiency Virus (FIV), the Simian Immunodeficiency Virus (SIV), and the Human Immunodeficiency Virus (HIV) - based on their respective hosts (these being ovine / caprine and bovine, equine, feline, and primate).

Figure 2. The Retroviridae



The molecular biology of HIV.

In order to understand the significance of retroviral proteases in the lifecycles of HIV and SIV it is helpful to have knowledge of the viral biology. In this section the molecular and structural biology of HIV is outlined. HIV is described, but a general appreciation of both HIV and SIV biology can be obtained, since they are essentially very similar.

In general the retroviral lifecycle (fig. 3) can be broken down into the following five stages:-

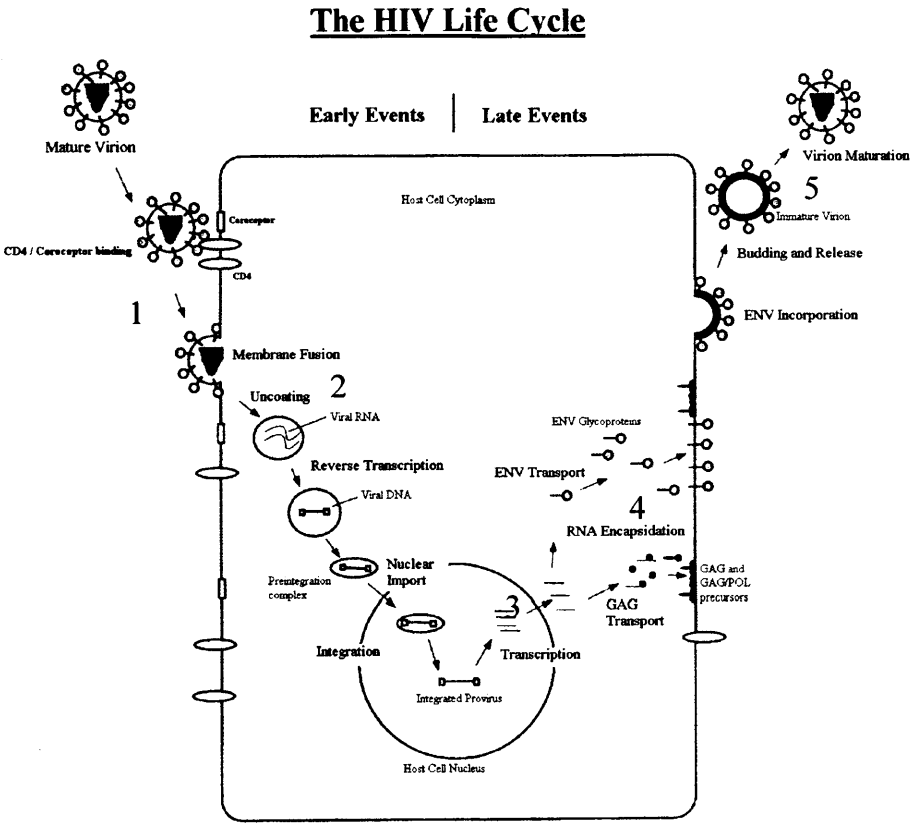
1. Fusion and entry.
2. Proviral synthesis and integration.
3. Proviral transcription and translation.
4. Assembly.
5. Budding and maturation.

Stages 1 and 2 comprise the early phase, while stages 3,4 and 5 make up the late phase. The stages of the lifecycle will be considered in detail, but first the structure of the mature virion will be briefly described.

Figure 3. The retroviral life cycle.

The retroviral lifecycle comprises five stages:-

1. Fusion and entry.
2. Proviral synthesis and integration.
3. Proviral transcription and translation.
4. Assembly.
5. Budding and maturation.



Structure of the mature virion.

The mature virion (Fig. 4) has a lipid bi-layer envelope, approximately 100 nm in diameter, forming an icosohedral, sphere-like structure. The lipid envelope is acquired during budding of the virion from the host cell. Since it was once part of the host cell membrane it contains a number of host cell membrane proteins. These include major histocompatibility antigens, actin, and ubiquitin (Arthur, Bess *et al.* 1992). In addition numerous 'spikes' are visible in electron micrographs protruding from the membrane surface. Each consists of two noncovalently associated envelope glycoprotein (gp) subunits - the 345 amino acid, 41 kilo Dalton (kD) transmembrane glycoprotein gp41 (or TM) which provides an anchor for the 515 amino acid, 120 kD surface glycoprotein, gp120 (or SU). The gp120 and gp41 subunits are encoded by the HIV env gene, and are initially synthesised as a precursor called gp160. Following its synthesis gp160 is directed into the host cell secretory pathway for transport to the cell surface. Whilst within the secretory pathway it is modified by glycosylation, and proteolytically processed into the mature gp120 and gp41 subunits (SU and TM) (Willey, Bonifacino *et al.* 1988; Earl, Moss *et al.* 1991; Haseltine 1991). The host cell protease responsible for this cleaveage has not yet been identified, however, if the processing event is prevented the viral surface glycoprotein is incapable of mediating membrane fusion (Kowalski, Potz *et al.* 1987; McCune, Rabin *et al.* 1988; Haseltine 1991) making it a target for antiviral therapy. Both envelope glycoproteins are involved in entry of the host cell, gp120 being especially important as it interacts with the $CD4^+$ receptor carried by T_4 cells (Lifson, Feinberg *et al.* 1986; Schnittman, Lane *et al.* 1988).

Lining the inner surface of the lipid membrane are 2000 copies[2] of the matrix protein p17 (or MA). Within the matrix, at the centre of the particle is the capsid - constructed from 2000 copies[2] of the p24 (or CA) capsid protein.

The capsid encases two molecules of the unspliced viral genomic RNA (HIV is diploid), which is stabilised by approximately 2000 copies[2] of the nucleocapsid protein p7 (or NC). Also within the capsid are the virally encoded enzymes required for replication - reverse transcriptase (RT), integrase (IN), and protease (PR) - plus three auxiliary proteins - NEF, VIF, and VPR. There are three additional auxiliary

[2] Estimate based on calculations by Forster, Mulloy and Nermut 2000.

proteins encoded by the genome - REV, TAT, and VPU - but these are apparently not packaged within the virion. The presence of the six auxiliary proteins is a characteristic of both the HIV and SIV groups, distinguishing them from the other retroviruses. VPU is found exclusively in HIV-1 and no homologues have been described in other primate lentiviruses (Schubert, Bour, *et al.* 1996) other than the HIV-1 related SIV_{CPZ} isolated from chimpanzees (Huet, Cheynier, *et al.* 1990). For a review of HIV-1 related accessory proteins see Emerman and Malim (Emerman and Malim 1998). An additional auxiliary protein, VPX, is absent in HIV-1 but present in HIV-2 and members of the SIV group. The VPR and VPX proteins are homologous. It is probable that the vpx gene within HIV-2 arose by duplication of an ancestral vpr gene. Evolutionary distance analysis has shown that both genes were well conserved when compared with viral regulatory genes, and indicated that the duplication occurred at approximately the same time as the HIV-2 group and the other primate lentivirus groups diverged from a common ancestor (Tristem, Marshall, *et al.* 1992).

The integrated HIV proviral genome (Fig. 5) is double stranded DNA, 10 kilobases (kb) long, consisting of two flanking 'Long Terminal Repeat' regions (LTR^s)[3] plus three main coding sequences, gag, pol, and env. The LTR^s contain regulatory sequences for HIV replication, whilst gag, pol and env encode the core proteins, the viral enzymes (eg. reverse transcriptase, protease and integrase), and the envelope glycoproteins (gp120 and gp41) respectively.

[3] Discussed in section entitled Proviral synthesis and integration, this chapter.

Figure 4. The mature virion.

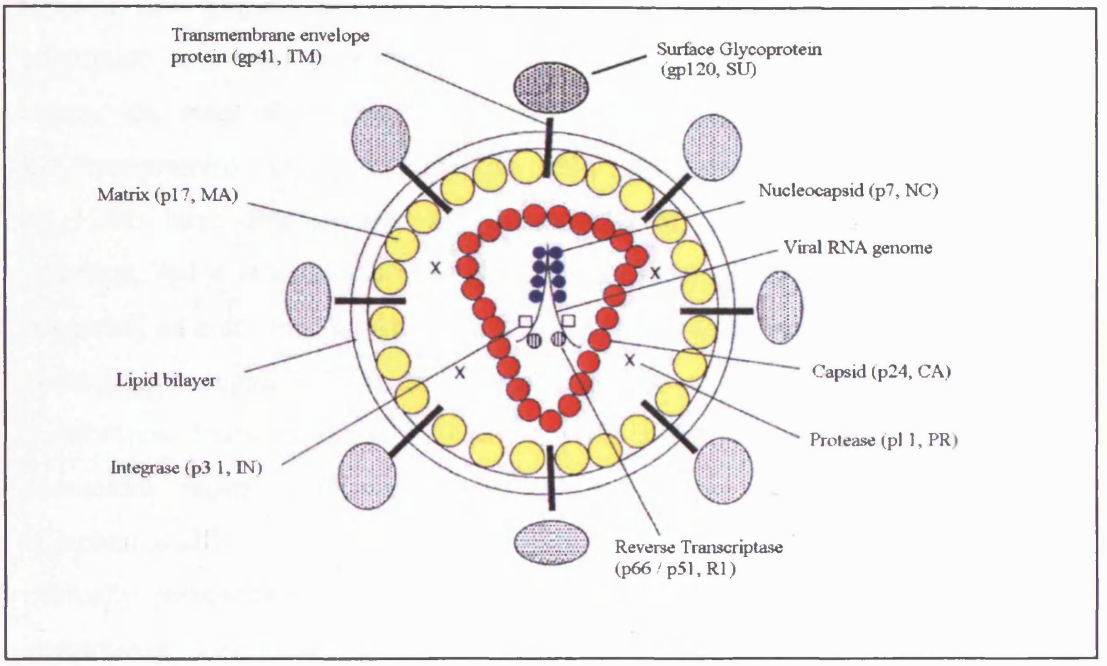


Figure 5. The HIV genome.

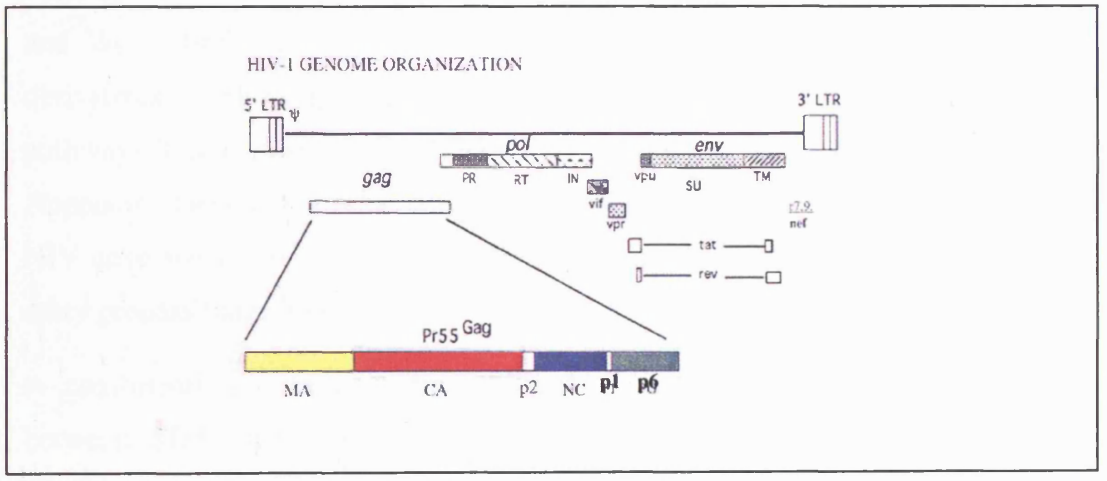


Figure 4 (above) illustrates the basic structural organisation of the mature virion, with the host derived lipid bilayer embedded with gp41/gp120 complexes and surrounding the matrix and capsid structures. Matrix, capsid and nucleocapsid proteins are coloured yellow, red and blue respectively, as are the corresponding genes encoding them shown in figure 5 (above) – a genomic organisation schematic for HIV-1. HIV-2 and SIV are essentially the same, but incorporate an additional auxiliary protein, VPX, which is homologous to VPR (which they also possess). VPU is found exclusively in HIV-1.

Fusion and entry.

Once a viral particle has entered the body and encountered a potential host cell, adsorption and entry may occur. Cells (predominantly T4 lymphocytes, but also monocytes, macrophages and some neural cells) expressing the $CD4^+$ antigen (a 60 kD glycoprotein) are all potential hosts. Dalgleish *et al.*(1984) (Dalgleish, Beverley *et al.* 1984) have demonstrated the dependance of HIV on the $CD4^+$ antigen for infection, and it is known to be a high affinity receptor for gp120 - their interaction triggering an entry mechanism. The sequential interaction of gp120, with CD4 and a chemokine receptor on the cell surface initiates fusion of the viral and cellular membranes. Parts of the gp120 bind to the $CD4^+$ receptor, and to a group of chemokine receptors (Kwong, Wyatt *et al.* 1998) that are essential cofactors (Clapham and Weiss 1997). There are seven of these G-coupled proteins, and they are normally responsible for mobilising intracellular calcium and inducing leukocyte chemotaxis. Viral entry is not possible in their absence thus their physiological ligands are able to inhibit viral entry by competing with HIV for their receptor (Bleul, Farzan *et al.* 1996; Choe, Farzan *et al.* 1996; Oberlin, Amara *et al.* 1996; Clapham and Weiss 1997). It may therefor prove possible to block infection using ligand derivatives which bind these receptors but do not activate the chemokine signalling pathways (Clapham and Weiss 1997). These interactions are considered further in the Appendix (General Introduction – supplementary information; "The interaction of HIV envelope glycoproteins with CD4 and Chemokine receptors facilitates the viral entry process" page 328).

A conformational change within the gp41 subunit is illicited by the interaction between $CD4^+$ and gp120. Each gp41 contains two 'heptad repeat' sequences (consensus motifs predicted to be hydrophobic alpha-helices), and crystallographic studies have shown that these two heptad repeat domains form a helical bundle containing a trimer of each domain (Chan, Fass *et al.* 1997; Tan, Liu *et al.* 1997; Weissenhorn, Dessen *et al.* 1997). It is suggested that these bundles have a role in the conformational changes essential for membrane fusion (Dubay, Roberts *et al.* 1992; Wild, Dubay *et al.* 1994). It may thus be that the HIV fusion mechanism is similar to the 'spring-loaded' mechanism proposed for the influenza virus whereby hemaglutinin changes from a loop to an extended coiled-coil and in the process moves a 'fusion peptide' into a position such that membrane fusion can occur (Carr and Kim

1993; Bullough, Hughson *et al.* 1994). Compounds able to prevent the fusion process have been discovered based on the C and N-terminals of the central trimer (Wild, Oas *et al.* 1992; Jiang, Lin *et al.* 1993; Wild, Dubay *et al.* 1994; Wild, Shugars *et al.* 1994; Lu, Blacklow *et al.* 1995) and it is probable that these act by interaction with the gp41 helical bundle preventing it from adopting its fusogenic conformation (Judice, Tom *et al.* 1997; Tan, Liu *et al.* 1997; Shu, Liu *et al.* 2000). Jiang and Debnath have reviewed the development of inhibitors targeting gp41 (Jiang and Debnath 2000).

Fusion of the viral and cellular membranes results in entry of the viral core into the host cell cytoplasm. Capsid breakdown then allows viral enzymes access to their respective substrates. It also releases the viral RNA genome, and this must then be reverse transcribed prior to its integration into the host cell genome.

Proviral synthesis and integration

Reverse Transcriptase, RT, (an RNA dependant DNA polymerase, characteristic of retroviruses) uses the viral RNA as a template for the synthesis of double stranded 'proviral' DNA. RT is present in HIV as a heterodimer with its two polypeptides, p66 and p51, being 66 kD (560 residues) and 51 kD (440 residues) respectively. These share a common N-terminus, and the 51 kD form appears to be derived from the 66 kD form by proteolytic cleavage of approximately 130 amino acids from the C-terminus. Both have an N-terminal polymerase domain consisting of four subunits (called fingers, palm, thumb, and connection). Both the 66 kD form, and a 15 kD polypeptide derived from its C-terminus display an RNase H activity. There is a 'tether region' in between the polymerase and RNase H domains of the 66 kD polypeptide.

Reverse transcription requires that a lysine tRNA primer anneals to a primer binding site near the 5' end of the viral RNA genome (Oude Essink, Das *et al.* 1996). The lysine tRNA primer is incorporated into virions during their assembly (Oude Essink, Das *et al.* 1996; Huang, Wang *et al.* 1997) and can be extended by several nucleotides while within the particle (Huang, Wang *et al.* 1997), however, the majority of reverse transcription probably occurs in the host cytoplasm following uncoating. RT moves from 3' to 5', each RNA base enters its binding site, and a complementary nucleotide is incorporated into the growing DNA chain that sits in the enzyme's primer binding groove. The result is a hybrid DNA / RNA molecule. The RNase H activity derived

from the p66 C-terminus digests the RNA strand of the hybrid DNA / RNA molecule. Cellular DNA polymerases then replace the RNA strand with a second DNA strand.

Since inception of this project the structure of RT has been solved (Kohlstaedt, Wang *et al.* 1992; Jacobo-Molina, Ding *et al.* 1993; Esnouf, Ren *et al.* 1995; Ren, Esnouf *et al.* 1995; Ren, Esnouf *et al.* 1995; Rodgers, Gamblin *et al.* 1995; Hsiou, Ding *et al.* 1996). During synthesis of the linear, double stranded DNA 'provirus' the Long Terminal Repeat (LTR) sequences flanking the coding regions of the genome are formed. Retroviral RNA has a sequence of 30-60 base pairs (bp) designated 'R' (redundant) at the extreme 5' and 3' ends. Each of the two R regions has an adjacent 'U' (unique) sequence, these are U5' and U3'. The U5' (80-120 bp) precedes the gag gene and the U3' (200-1200 bp) follows the env gene (see fig. 5). During DNA production a series of template switches by RT repeats the U3'-R-U5' sequence at each end of the new DNA strand forming the LTR's. Retroviral LTR's are 280-1300 bases long depending on the virus - in HIV each LTR is 634 bp long. They are essential for integration - the process whereby viral DNA is translocated into the host cell nucleus and incorporated into its genome. They allow circularisation of viral DNA, giving two types of covalently closed, supercoiled molecules. The first type incorporates one LTR, the second two tandem copies. The second type is thought to be the precursor to the integrated provirus, and is probably formed by ligation of the blunt ends of the linear molecule. A novel sequence at the junction between the two LTR's called the 'circle junction' is probably the substrate for the viral endonuclease, 'integrase' (IN). This enzyme is encoded by the 3' end of the pol gene and seems to cut both the host and viral DNA, giving staggered ends. Probably the short end of the viral DNA covalently joins to the protruding end of the host DNA, whilst DNA repair synthesis fills in the host DNA and removes the non-hydrogen bonded terminal viral dinucleotide sequences. This mechanism explains the observed deletion of LTR terminal dinucleotide sequences and the short duplication of host DNA flanking the integrated provirus. If the terminal dinucleotide sequences (TT/AA) are recognition sites for integration, their loss renders the process irreversible, explaining the stability of the integrated provirus. The site of integration within the host genome appears to be random, though DNA exposed during replication, actively transcribed regions, nicks, gaps and other modifications have been suggested as possible targets. A preference for kinked or distorted DNA has been shown *in vitro* (Miller, Bor *et al.*

1995), and this may indicate the involvement of nucleosomes as integration sites. It may be that interaction with cellular DNA binding proteins guides IN to specific sites. Much of the viral DNA remains unintegrated in the cytoplasm - an unusual characteristic of HIV, and one that may be a factor in its cytopathic effect.

It has been observed that there are similarities between retroviral proviruses and transposons (Finnegan 1981; Hehlmann, Brack-Werner *et al.* 1988), they are structurally related (Majors, Swanstrom *et al.* 1981), and it is even possible that retroviruses evolved from transposons (Flavell 1981; Shimotohno and Temin 1981; Finnegan 1983; 1992). Models for the molecular aspects of retroviral integration have been based on knowledge of transposases (Polard and Chandler 1995; Capy, Vitalis *et al.* 1996; Haren, Ton-Hoang *et al.* 1999), which are structurally similar to retroviral integrases (Rice, Craigie *et al.* 1996).

IN is packaged within the viral particle as a Gag-Pol precursor. PR releases the active IN from the precursor by proteolysis, either during or after the budding process. It has 288 residues, three functional domains - the N-terminal zinc finger-like domain, the central catalytic domain, and the C-terminal domain. NMR structures of the N and C terminal domains were solved individually (Lodi, Ernst *et al.* 1995; Cai, Zheng *et al.* 1997; Cai, Huang *et al.* 1998). However, the catalytic domain exhibited poor solubility, limiting any structural analysis until a more soluble mutant (having a single change – F185K) was made which enabled a crystal structure to be obtained (Dyda, Hickman *et al.* 1994).

As yet the structure of IN in its entirety has not been established, neither is the precise mode of IN action clear. The arrangement and interaction of the three domains is not yet known, but the fact that each domain has been shown to exist individually in a dimeric form suggests that this may be true of IN as a whole. However, there has been some suggestion that a tetramer is more likely (Rice, Craigie *et al.* 1996; Cai, Zheng *et al.* 1997).

Proviral transcription and translation

A period of dormancy (latency) may follow integration, until a signal (eg. T-cell activation) causes transcription of the viral DNA via a promoter located in the 5' LTR. In addition to the HIV promoter the 5' LTR contains the regulatory elements for RNA polymerase II transcription, and upstream of the start are sites for the binding of cellular transcription factors including Nuclear Factor-kB (NF-kB), Specificity protein 1 (Sp1) and TATA-box binding protein (TBP). These factors may help to control the rate of transcription initiation thus their relative abundance in the host cell may influence the level of proviral replicative activity. *In vitro* phytohemaglutanin (PHA), phorbal myristate acetate (PMA), and other mitogens, activate T-cells - probably via cell surface receptors, protein kinase C, and the secondary messenger inositol triphosphate. In this way they mimic the action of an antigen, and stimulate the production of the cellular transcription. These in turn induce the production of interleukins and their receptors, but also increase LTR directed viral gene expression by binding the upstream sites within the LTR. Deletion studies have shown that tandemly repeated core elements of the LTR's, function as transcription enhancers (expression of the auxiliary proteins TAT and REV is enhanced in this way). Their activation ends the proviral latency, and the first viral mRNAs produced are doubly spliced and encode TAT, REV and NEF.

Once transcription is underway a full length RNA transcript of approximately 9000 bp is produced. Nine important regions have been identified within the full-length transcript (see fig. 5).

However, even in the presence of the cellular factors described, transcription complexes initiated at the 5' LTR are inefficiently elongated. This is overcome by the action of the tat (transactivator) gene, which has a positive effect on both its own expression and that of other genes by increasing the ability of transcribing polymerases to elongate their transcripts. LTR directed gene expression results in production of another transactivator called REV (Regulation of Expression of Viral proteins). As REV levels increase it has the effect of reducing the frequency of splicing so that singly spliced or unspliced transcripts are now transported to the cytoplasm so that translation of other viral components can begin, and so that full length genomic RNA becomes available for packaging. TAT and REV are examined in more detail below.

TAT

The tat gene encodes a 14 kD (86 residue) protein (TAT) found in the nucleus of infected cells. TAT may recruit or activate factors that phosphorylate the RNA polymerase II C-terminal domain (CTD), and enhanced phosphorylation of this domain is known to occur as the polymerase II changes function from initiation to elongation (Parada and Roeder 1996; Yang, Herrmann *et al.* 1996; Zhou and Sharp 1996; Cujec, Cho *et al.* 1997; Garcia-Martinez, Mavankal *et al.* 1997; Jones 1997). In this way TAT increases the production of viral mRNA by about 100 fold. In the absence of TAT the polymerase fails to transcribe beyond a few hundred nucleotides, presumably because the CTD remains unphosphorylated. Consequently TAT is essential for replication and is another target for antivirals. The combined effects of TAT and T-cell activation increase LTR directed gene expression by 800 times the basal level seen in non-activated cells where tat is not expressed.

Typically transactivators bind DNA sites, however, TAT binds to an RNA hairpin loop called TAR (Trans-Activating Response element). TAR is located at the 5' end of developing viral transcripts (ie. in the R region of the LTR encoding the 5' leader of viral mRNA). An arginine rich region of TAT allows binding to a tri-nucleotide bulge in TAR (Bayer, Kraft *et al.* 1995). TAT protein may stabilise RNA molecules that initiate with the LTR TAR sequences. Alternatively it may have a transcription anti-termination activity, since TAR contains a 24 nucleotide inverted repeat - potentially able to form a stable hairpin structure in the mRNA leader and thus terminate transcription. Whatever TAT's mechanism of action it probably regulates viral protein expression post-transcriptionally. It may facilitate transport of TAR-containing RNA molecules from a compartment in the nucleus (the nucleolus) within which RNA is made and spliced, to a subcellular compartment where it is accessible to the translational initiation complex. The overall effect of the TAT-TAR positive regulatory loop is explosive replication of the virus under favourable conditions. The TAT-TAR interaction is stabilised by an as yet unidentified host cell protein, which may be encoded by chromosome 12 (Alonso, Cujec *et al.* 1994).

REV

The 19 kD REV protein is encoded by sequences overlapping those of TAT but which are in a different reading frame. REV is essential for the formation of the long HIV transcripts encoding GAG, POL and ENV precursor polypeptides. In cells infected by a rev defective HIV genome, levels of GAG, POL and ENV mRNA are depressed, whilst normal overall levels of viral mRNA are maintained by a compensatory increase in short, doubly spliced species. This suggests that REV acts by relieving a block in GAG and by allowing the formation, but not splicing of, gag pol and env precursor mRNA transcripts. Since all viral proteins are cleaved from precursor polyproteins encoded by such full-length transcripts, the rev gene is a positive regulator of these proteins.

Although REV may simply inhibit the splicing process it appears more likely that it targets the nuclear export system. In doing so it may steer the pre-cursor mRNA away from the splicing and degradation pathways (its usual fate) allowing its transport from the nucleus to the cytoplasm prior to complete splicing (Hope 1997).

Following REV's activity spliced and genomic length viral mRNA is transported to the host cell cytoplasm where its translation results in the GAG and GAG-POL polyproteins. These subsequently become localised to the host cell membrane and are later processed into their constituent proteins (MA, CA, NC, p6, PR, RT and IN).

Assembly

In 1968 Maizel *et al.* and Summers *et al.* (Maizel and Summers 1968; Summers and Maizel 1968) showed that in the poliovirus the combined molecular weights of protein comprising the virion exceeded the coding capacity of its genome. This lead to the realisation that the viral proteins were expressed as a single large precursor polypeptide (Holland and Kiehn 1968; Jacobson and Baltimore 1968; Summers and Maizel 1968). It was Jacobson and Baltimore who termed the precursor a polyprotein (Jacobson and Baltimore 1968). Polyproteins are made by all retroviruses including HIV and SIV, and many plus strand RNA viruses. They contain many distinct domains which are cleaved proteolytically, either as they are translated (RNA viruses) or during virion maturation (Retroviruses).

Many viruses have evolved such that they produce more structural proteins than non-structural proteins simply because they need more of the former type. Retroviruses produce more GAG and ENV proteins than POL proteins. The ENV proteins are derived from subgenomic mRNA species produced by splicing. The GAG polyprotein is translated from genome length mRNA, while the POL proteins are also encoded by genome length mRNA but translated as a GAG-POL polyprotein. The level of GAG-POL polyprotein production is 80-95% lower than that of the GAG polyprotein. This is due to two rare translational events – the suppression of an amber codon (at end of gag coding sequence), or frame shifting at the 3' terminal half of the gag coding sequence.

Localisation of the GAG and GAG-POL polyproteins to the inner surface of the host cell membrane is made possible by the matrix protein MA, located at the N-terminal end of the GAG polyprotein. This is an initial stage of the assembly process. At this point some of the individual components of the polyproteins will be considered, including MA, CA, NC and p6. As previously stated, the enzymes PR, RT and IN are incorporated within the GAG-POL polyprotein, and thus also become attached to the host cell membrane. However, their involvement in the lifecycle will be considered later (see Budding and Maturation, this Chapter page 46).

MA

MA has an N-terminal myristate group and a number of basic residues (lysines 26, 27 30 and 32) among the first 50 amino acids of its N-terminal domain. Both are involved in an interaction with the host cell membrane. Crystallographic studies have revealed that MA exists as a trimer with the basic residues of each monomer positioned at what is presumably the membrane-binding surface. It is probable that the trimeric structure is of biological significance since mutagenesis of residues involved in trimerisation (amino acids 42-77) prevents viral assembly. Probably the three myristates penetrate the host cell's lipid bilayer, while the lysines from each monomer interact with the phospholipid head groups at the cytoplasmic surface of the membrane. In addition to locating GAG and GAG-POL polyproteins to the host cell membrane, MA may also ensure the incorporation of ENV glycoproteins with long cytoplasmic tails into viral particles (Freed, Englund *et al.* 1995; Mammano, Kondo *et al.* 1995; Hill, Worthylake *et al.* 1996; Massiah, Worthylake *et al.* 1996).

CA

The other components of the polyproteins are suspended from the membrane beneath MA. The first of these is the capsid protein CA, approximately 2000 copies of which will eventually form the viral core. CA subunits have been seen to arrange themselves in strips within a crystal, possibly a reflection of their packaging arrangement *in vivo*.

CA is 231 amino acids long and has two domains - the N-terminal (residues 1-151), and the C-terminal (residues 152-231). The structures of both these domains have been solved either by crystallography or NMR (Gamble, Vajdos *et al.* 1996; Gitti, Lee *et al.* 1996; Momany, Kovari *et al.* 1996; Gamble, Yoo *et al.* 1997). Mutations of the N-terminal domain do not disrupt assembly or budding, but do reduce viral infectivity. It may be that this domain is involved in uncoating, and it is known to associate with Cyclophilin A (CypA) - probably a cellular chaperone. The N-terminal domain exists as a monomer in solution.

The C-terminal domain has an extended strand followed by four alpha helices and exists as a dimer in solution. It has a 20 amino acid region that is highly conserved throughout retroviral GAG proteins called the Major Homology Region (MHR). This region produces a tight fold stabilised by hydrogen bonds between its four most conserved residues. It is essential for particle assemby (Srinivasakumar,

Hammarskjold *et al.* 1995) and has a possible role in membrane affinity (Ebbets-Reed, Scarlata *et al.* 1996).

NC

The third component of the GAG polyprotein is the nucleocapsid protein NC, which following maturation is responsible for coating the genomic RNA inside the virion core. In coating the RNA it is able to protect it from ribonucleases, and ensure that the genome is compact enough for packaging into the core. NC has additional chaperone like functions and is involved in promoting $tRNA^{Lys}$ primer annealing, melting of RNA secondary structure, DNA strand exchange during reverse transcription (Cameron, Ghosh *et al.* 1997; Guo, Henderson *et al.* 1997; Huang, Barchi *et al.* 1997), and stimulation of integration (Carteau, Batson *et al.* 1997).

NC is a basic protein of 55 residues and it's structure has been determined by NMR (Morellet, Jullian *et al.* 1992; Summers, Henderson *et al.* 1992). It has two zinc finger domains (CCHC type zinc fingers), and removal of the zinc by disulphide substituted benzamide compounds inhibits viral replication. NC binds non-specifically to single stranded nucleic acids, but also binds specifically to the packaging signal ψ, thus ensuring that full length genomic RNA becomes incorported into the virion. The zinc fingers and the basic residues flanking them are required for the interaction with ψ (Poon, Wu *et al.* 1996; Schmalzbauer, Strack *et al.* 1996). The packaging signal ψ probably consists of 3 RNA hairpin loops surrounding a major splice donor site (Clever and Parslow 1997; Laughrea, Jette *et al.* 1997). The first of these hairpins includes the 'kissing loop' responsible for RNA dimerisation.

p6 and VPR / VPX

Viral protein R (VPR) is an auxiliary protein that is conserved in both HIV-1 and HIV-2 (Dedera, Vander Heyden, *et al.* 1989), it is a 96 amino acid, 15 kD protein that increases the rate of replication and accelerates the cytopathic effect of the virus in T-cells by acting in trans to increase levels of viral protein expression (Cohen, Terwilliger, *et al.* 1990). The C-terminal 51 amino acids of GAG make up p6, a protein which binds VPR and is involved with the incorporation of VPR during viral assembly (Checroune, Yao *et al.* 1995; Lu, Bennett *et al.* 1995; Kondo and Gottlinger 1996) - an alpha-helix at the N-terminal of VPR contains amino acids responsible for binding p6. In addition p6 incorporates a sequence of four amino acids (Pro7 Thr8

Ala9 Pro10) involved in particle release (Huang, Orenstein *et al.* 1995). Viral protein X (VPX) is a 16 kDa accessory protein, homologous with VPR, expressed in cells infected with HIV-2 and most SIV strains (Horton, Spearman *et al.* 1994). In HIV-2 and SIV VPX is necessary for efficient replication in Peripheral Blood Mononuclear Cells (PBMC) (Park and Sodroski 1995).

The ENV polyprotein (gp160), VPU, and NEF

Following its translation the env polyprotein (gp160) undergoes glycosylation within the Endoplasmic Reticulum (ER), followed by cleavage into SU (gp120) and TM (gp41). In preparation for budding gp41 and gp120 must be incorporated into the host lipid membrane, creating the outer coat for a new virion. However, the host $CD4^+$ receptors are also glycosylated in the ER - consequently the gp160 or gp120 and CD4+ interact prematurely to form unwanted complexes. These complexes can inhibit the formation of the fully functional gp41/gp120 complex and its translocation to the cell membrane. The gp160/gp120 must therefore be released from these complexes. Auxliary proteins VPU and NEF assist in degradation of $CD4^+$ throughout the host cell.

VPU

The auxiliary protein VPU (found only in HIV-1 group virions) assists in the release by promoting degradation of $CD4^+$. The degradation of $CD4^+$ is interupted by proteasome inhibitors such as lactacystin (Fujita, Omura *et al.* 1997), and is regulated by a cellular kinase - which phosphorylates Ser52 and Ser56 of VPU. If these residues are mutated a decrease in CD4+ degradation is observed (Cohen, Subbramanian *et al.* 1996). It has now been shown that VPU interacts with an F box protein called β-Transducin repeat Containing Protein (β-TrCP) in a phosphorylation dependent manner (Margottin, Bour *et al.* 1998; Akari, Bour *et al.* 2001; Coadou, Gharbi-Benarous *et al.* 2003). In doing so VPU is able to corrupt the ubiquitin pathway and directs the β-TrCP to the host's own $CD4^+$ protein - this is described more fully by Yaron *et al.* (Yaron, Hatzubai *et al.* 1998). VPU is somehow also able to reduce the occurrence of Major Histocompatibility Complex class I (MHCI) proteins on the cell surface. This probably prevents cytotoxic T-lymphocytes from recognising that the cell has been infected and thus killing it (Kerkau, Bacik *et al.* 1997).

VPU is an oligomeric integral membrane phosphoprotein of 81 amino acids (Cohen, Terwilliger, *et al.* 1988; Strebel, Klimkait, *et al.* 1988; Strebel, Klimkait, *et al.* 1989), with a transmembrane N-terminal domain (residues 1-24) and a C-terminal cytoplasmic tail (residues 25-81). It is probable that the C-terminal tail is involved in receptor binding and $CD4^+$ degradation (Willey, Maldarelli *et al.* 1992; Schubert and Strebel 1994), and that located within it are the (serine) phosphorylation sites regulating these functions. The N-terminal domain is probably involved in the release of particles from the host cell, and also displays non-specific cation channel activity (Strebel, Klimkait, *et al.* 1988; Strebel, Klimkait, *et al.* 1989; Terwilliger, Cohen, *et al.* 1989; Schubert, Bour, *et al.* 1996; Ewart, Sutherland *et al.* 1996). Mutation of this domain results in virions with an increased tendancy to remain associated with the cell surface, or localised to intracellular membranes.

Although VPU is not essential for virus replication *in vitro*, its expression enhances viral particle release from the host cell in tissue culture systems (Strebel, Klimkait, *et al.* 1988; Strebel, Klimkait, *et al.* 1989; Terwilliger, Cohen, *et al.* 1989). In its absence viral proteins accumulate within the infected cell as a result of intracellular budding, causing increased cytopathicity in vpu-deficient strains (Klimkait, Strebel, *et al.* 1990). Given the absence of vpu from the non-HIV-1 type retroviruses they were envisaged to have evolved alternative mechanisms to augment particle release. Studies by Bour, Strebel and co-workers have shown that the gp140 envelope protein of HIV-2_{ROD} has VPU-like activity and can enhance particle release (Bour, Schubert, *et al.* 1996; Bour and Strebel 1996). There are similarities between VPU and gp140 – both are integral membrane proteins able to form oligomeric structures, they are similarly regulated (since in HIV-1 ENV and VPU are translated from the same bicistronic mRNA), and it is probable that they both act in similar subcellular compartments (Bour, Schubert, *et al.* 1996 and references therein). It seems then that in the viral groups where VPU is absent its function is carried out by the ENV protein gp140.

NEF

NEF is an N-terminally myristoylated protein of 206 amino acids, and promotes the endocytosis and degradation of $CD4^+$ receptors already at the cell surface. The removal of CD4+ molecules from the surface may be a means of avoiding an immune response, but would also be necessary to ensure that no $CD4^+$ receptors will be present on the viral outer envelope. Such an occurence would presumably result in viral particles attempting to fuse with each other. It probably also helps prevent multiple infection of a host cell.

NEF probably interacts with the tail of $CD4^+$ (Mangasarian and Trono 1997), presumably via its myristoylated N-terminus. NEF then directs the $CD4^+$ from its location at the cell surface or at the golgi, to lysosomes where receptor degradation occurs (Mangasarian and Trono 1997) NEF also shares VPU's ability to downregulate the deployment of MHC class I proteins at the cell surface (Le Gall, Heard *et al.* 1997).

Approximately 70 molecules of NEF are incorporated into each new virion, and these are cleaved by PR (at residue 57) during maturation to yield a soluble C-terminal fragment (Guatelli 1997). This suggests possible involvement in assembly, maturation and entry.

At this stage in development (and as previously described) the GAG polyprotein is being synthesised in the ribosomes from unspliced mRNA. A smaller amount of GAG-POL precusor proteins are also expressed following a translational frameshift. The N-terminal myristoylated MA domain of the GAG and GAG-POL polyproteins, besides enabling them to bind directly to the inner surface of the host cell membrane, also allows interaction with the cytoplasmic tail of gp41.

Budding and maturation.

Gag and gag-pol polyproteins migrate to the host cell membrane where they become anchored via an N-terminal myristic acid group (Henderson, Krutzsch *et al.* 1983; Mervis, Ahmad *et al.* 1988). Between 1200 and 2000 such polyproteins per new particle bind the inner surface of the host membrane. The adjacently anchored polyproteins then interact with each other, and electron micrograph observations show them apparently aggregating together (Gonda, Wong-Staal *et al.* 1985). Presumably this is at least in part the result of dimer formation between adjacent monomers of each component protein. This interaction distorts the host cell membrane, forming a bulge. The bulging effect becomes more exaggerated as more polyproteins become anchored in the vaccinity, and join the aggregate. Eventually the bulge is an almost spherical protruberance from the host cell surface, and finally the sphere is completed as the host cell membrane ruptures at the point of maximum distortion. The spherical virion buds off from the host cell's plasma membrane, encapsidating two full-length molecules of the unspliced genomic RNA in the process.

Within the newly detached virion the polyproteins are still intact, none of their component proteins have been released, the particle is immature and not infectious. In order for maturation to occur the active PR must be present (Kohl, Emini *et al.* 1988). As adjacent, membrane anchored, viral polyproteins associate or aggregate their component proteins may be able to dimerise into their active conformations. Among these emerging dimers is the PR, which on becoming an active dimer is now able to recognise, bind and cleave the scissile sites between component proteins of adjacently aggregated polyproteins. There have been several accounts suggesting that once dimerised PR is able to autocatalytically cleave itself from within its gag-pol polyprotein (Farmerie, Loeb *et al.* 1987) (Debouck, Gorniak *et al.* 1987; Giam and Boros 1988) (Krausslich and Wimmer 1988). The first five N-terminal residues of the PR (linking PR to the gag encoded half of the polyprotein) serve no known catalytic or structural function, and it may be that their sole purpose is to provide the flexible linker required to make autocatalysis possible. It has been suggested that these residues are sufficiently flexible that the N-terminal scissile site can be manouvered into its own active site for cleavage. Mutagenesis of residues at the N-terminal cleavage site has been successful in modifying autoproteolytic activity (Rose, Salto *et al.* 1993)

Once PR is released, either by itself or perhaps by its neighbour, it is able to cleave nearby polyproteins into their constituent protein subunits, releasing more PR in the process. A cascade of proteolysis consequently ensues, until all the polyproteins are processed, and maturation is complete.

The 23 kD protein VIF (Viral Infectivity Factor) is also required at this stage, although its mode of action is unclear. Mutants of vif are unable to initiate particle-mediated infection, though they are still able to spread cell-to-cell, and infected cells may still form syncitia (multi-nucleated giant cells produced by the fusion of infected host cells - expressing gp120 on their surface - with other $CD4^+$ carrying cells). Syncitia formation is a major cause of HIV induced cell death. The defect in the vif gene mutants is not related to virus adsorption, packaging of RNA, or reverse transcription. It may be that VIF is involved in post-translational modification of viral proteins.

It is the viral protease (PR), however, which plays the major role in maturation, and its activity yields the individual enzymes required later for replication (IN and RT), plus the major structural proteins (MA, CA, and NC). During the maturation process the structural proteins rearrange and an infectious particle results. If PR activity is blocked by mutation or inhibition the maturation process cannot proceed, and the viral particles remain non-infectious. Thus PR has long been recognised as a major antiviral target. It has consequently been the subject of intense investigation, culminating in the widespread clinical use of PR inhibitors today. The next section will deal with this enzyme in more detail, and will describe the investigations that have put it among the best-understood proteins known.

An introduction to proteases

Preliminary note

In 1992 when this PhD research project was begun the structure of HIV-1 PR apoenzyme had been solved, however, no structure existed for HIV-2 PR, or for an SIV PR (SIV is interesting as an animal model for HIV). Inhibitor design was predominantly based on non-scissile substrate analogues. No anti-PR drugs had yet been launched. The remainder of this introduction will consider the PR field as it developed up until 1992, and the questions that were being addressed by researchers at that time. Subsequently it will review developments contemporary with this study (1992 – 1996), and finally bring the reader up to date with the field. In this way the work that follows will be placed in context. At the end of the thesis the General Discussion will appraise the work with respect to the PR field as it is today, and assess the results in light of current knowledge.

Classification of proteolytic enzymes

Proteolytic enzymes are proteins able to catalyse the cleavage of peptide bonds. These enzymes engage either limited or unlimited proteolysis, the former involving the cleavage of only one or a limited number of specific peptide bonds within a target protein, whilst the latter involves the degradation of the target protein into its constituent amino acids. Ubiquitination of a protein is often the first step in a pathway involving ATP (Adenosine Tri-Phosphate) driven unlimited proteolysis, and a similar process of rapid protein degradation also occurs in the lysosome. Unlimited proteolysis is thus a means by which proteins may be removed from circulation and broken into their consituent amino acids. Limited proteolysis, by contrast, is often a means by which pro-enzymes may be activated, the cleavage of specifically recognised peptide bonds leading to release of an active mature protein from a formerly inactive precursor. The proteolytic enzymes (peptide bond hydrolases) have been classified by the International Union of Biochemistry and Molecular Biology (IUBMB) (1984) as peptidases (the term protease is synonymous with peptidase). Within the peptidase subclass there are two further groups of enzymes, the endopeptidases and the exopeptidases. These cleave bonds within a peptide chain in the case of endopeptidases, or sequentially remove residues from the N or C terminal ends of the chain in the case of exopeptidases. The term proteinase is synonymous with endopeptidase. This thesis deals with retroviral proteases engaging in limited

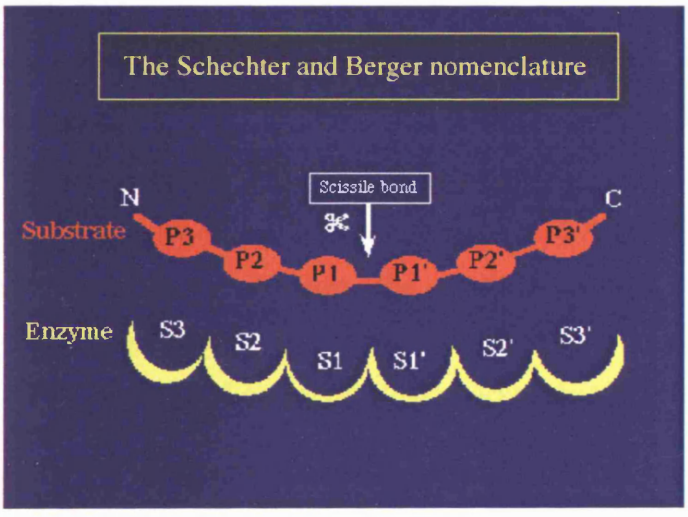
proteolysis of viral polyproteins at specifically recognised sites within the target protein. They are thus correctly classified as endopeptidases, or proteinases, whilst the name currently recommended by the Merops database is HIV retropepsin. However, the term protease is so widely applied to these enzymes in the literature that it has become almost universally accepted. Protease has therefore been used throughout this work in relation to the retroviral enzymes, and is also represented by the shorthand PR in that respect.

The endopeptidases / proteinases may be further divided according to their mechanisms of action. Four mechanistic classes have been recognised by the IUBMB, these being the serine, cysteine, aspartic and metallo proteinases. It is predominantly the aspartic proteinases that will be dealt with here, though some mention will be made of cysteine proteinases in chapter 5.

Protease subsite nomenclature

Where protease enzyme subsites or peptide substrates are described in this text the nomenclature of Schechter and Berger was used (Schechter and Berger 1967) whereby the subsites on the protease are called S and the substrate amino acid residues are called P (Fig. 6). Substrate peptide amino acid residues on the N-terminal side of the peptide bond are numbered P3, P2, P1, whilst those on the C-terminal side are P1', P2', P3' and so on, up to P8 and P8'. The residue numbers increase away from the scissile bond, so that the P1 and P1' residues are those flanking the bond. The corresponding subsites on the protease, interacting with the peptide residues of the substrate, are numbered S3, S2, S1, S1', S2', S3', in a similar way - thus complementing the nomenclature for the substrate.

Figure 6 The Schecter and Berger nomenclature.



The retroviral proteases (PR)

Following translation of viral proteins and the assembly of the viral core, the membrane coat incorporating SU and TM surrounds the particle, and the virion buds from the cell's surface and is released.

At this stage the particle is not infectious. A number of maturation events must take place in the period from assembly to budding, and the exact timing of these events is not yet clear. It had already been shown that in several other retroviruses a protease, usually encoded by the pol gene, was essential for this maturation process to occur (Dickson 1984). Indeed if the protease is absent or inactivated then the maturation process cannot be completed, and a non-infectious immature particle results. It was also known that the maturation process involves cleavage of the viral polyproteins into their constituent subunit proteins. In HIV it was shown by mutational analysis that the protease, PR, responsible for this cleavage process is located at the amino terminal coding region of the pol gene (Kramer, Schaber *et al.* 1986). It was clearly a potential target for antiviral therapy, especially since it was shown to be responsible for the release of structural proteins from the polyprotein encoded by gag, and for the release of viral enzymes from the GAG-POL fusion polyprotein (Kramer, Schaber *et al.* 1986; Debouck, Gorniak *et al.* 1987; Darke, Nutt *et al.* 1988). This suggested that the PR was able to release itself from the GAG-POL fusion polyprotein by autoproteolysis (Witte and Baltimore 1978; Crawford and Goff 1985; Katoh, Yoshinaka *et al.* 1985; Farmerie, Loeb *et al.* 1987; Darke, Nutt *et al.* 1988).

It was shown that the PR incorporated a single highly conserved triad (Asp-Thr-Gly) typical of the aspartyl proteases, and it was conjectured that it might belong to this family of enzymes (Toh, Kikuno *et al.* 1985; Power, Marx *et al.* 1986). Other members of this protease family had already been studied, in particular pepsin (produced in the stomach and involved in protein digestion), renin (found in blood plasma and involved, in association with angiotensin, in the regulation of blood pressure), cathepsin D (a lysosomal enzyme), rhizopuspepsin, endothiapepsin and penicillopepsin (isolated from fungi). For further information relating to renin see: Appendix; General Introduction – supplementary information; "The development of Renin inhibitors" (page 332).

Three dimensional crystal structures for endothiapepsin (Cooper, Foundling *et al.* 1987) and for penicillopepsin (Hsu, Delbaere *et al.* 1977) had already been solved.

However, these other mammalian and microbial aspartic proteases were typically in excess of 300 amino acids in length, and always had a pair of conserved triads - whereas the retroviral enzymes were much shorter and contained just one triad. Retroviral proteases were generally observed to have molecular weights in the range 11 – 14 kD, whereas the conventional aspartic proteases had molecular weights in the range 32 – 36 kD. HIV PR was eventually shown to consist of only 99 amino acids (Debouck, Gorniak *et al.* 1987; Graves, Lim *et al.* 1988; Nutt, Brady *et al.* 1988; Schneider and Kent 1988), making it the smallest of the retroviral proteases. A possible explanation for these discrepancies became apparent when the 11 kD MLV protease was shown to run at 22 kD on an analytical size exclusion column, suggesting that it was a dimer in solution (Yoshinaka and Luftig 1980). Perhaps this was true of all retroviral proteases?

In a conventional aspartic protease the two catalytic triads are brought into opposing proximity once the protein is correctly folded, forming the enzyme's active site. It was therefore proposed that the PR sequence represented the equivalent of a single domain of a conventional aspartic protease, and that furthermore it was a dimer in its active form (Pearl and Taylor 1987). In the dimeric state its catalytic residues would be brought together at the interface between the monomers and create an active site. Each monomer would contribute one of the two aspartates essential for catalytic activity. Pearl and Taylor performed extensive sequence alignments between members of the aspartic protease family, and the supposed retroviral aspartic proteases. This enabled them to make secondary structure predictions for the retroviral sequences, and compare them with the observed secondary structures for penicillopepsin and endothiapepsin. They found a close correspondence between the predictions and the observed structures, supporting the idea that the retroviral proteases were effectively equivalent to single aspartic protease domains. There is a difference in length between a retroviral protease and an aspartic domain (the former is typically about 20 amino acids shorter than the latter).

However, this can be explained in two ways: first by differences between the lengths of loops joining strands of ß-sheet in the respective proteases; and second by the fact that the retroviral proteases lack residues at their C-terminus that are present in other aspartic proteases.

Pearl and Taylor also constructed a three-dimensional model from the retroviral protease sequences based on the structure of endothiapepsin. They chose to model HIV-1 PR in particular, and took into account variations between the loop lengths of different enzymes by incorporating the loop lengths predicted for HIV-1 protease. The model showed that HIV-1 PR would essentially share the same folds as endothiapepsin. When modelled as a monomer the HIV-1 PR lacked the active site typical of an aspartic protease. However, a further model having two HIV-1 PR monomers, arranged opposite each other by using the twofold axis from the endothiapepsin active site, produced a plausible PR active site. In addition this model illustrated the presence of a substrate-binding cleft of sufficient size to accept as many as six amino acids.

It was later shown experimentally that the active HIV PR was dimeric (Darke, Leu *et al.* 1989) and that recombinant HIV PR was active in the dimeric form (Meek, Dayton *et al.* 1989). Subsequent crystallographic analyses finally confirmed the validity of the Pearl and Taylor models.

In forming its protease by the fusion of two identical monomers the virus is able to express an active protease while reducing the RNA required by half. In fact it is now clear that many viral proteins are dimeric suggesting that this is a widespread strategy for the reduction of genetic material to be packaged into a viral particle. A number of observations soon firmly established HIV PR as a member of the aspartic protease family. Inhibition of a protease by pepstatin (Morishima, Takita *et al.* 1970; Umezawa, Aoyagi *et al.* 1970) is indicative of aspartic proteases. When it was shown that pepstatin inhibited HIV-1 PR *in vivo* this suggested that the protease was a member of the aspartyl family (Hansen, Billich *et al.* 1988; Seelmeier, Schmidt *et al.* 1988; Darke, Leu *et al.* 1989). HIV PR is also inactivated by diazoacetyl norleucine methylester in the presence of Cu^{2+} - a reaction specific to aspartic proteases. In contrast, general protease inhibitors including leupeptin, chymotrypsin, phosphoramidone, and E-64 do not inhibit HIV PR even at 1 mM.

Figure 7 HIV-PR 3D structure, as published in Nature by Navia *et al.* 1989.

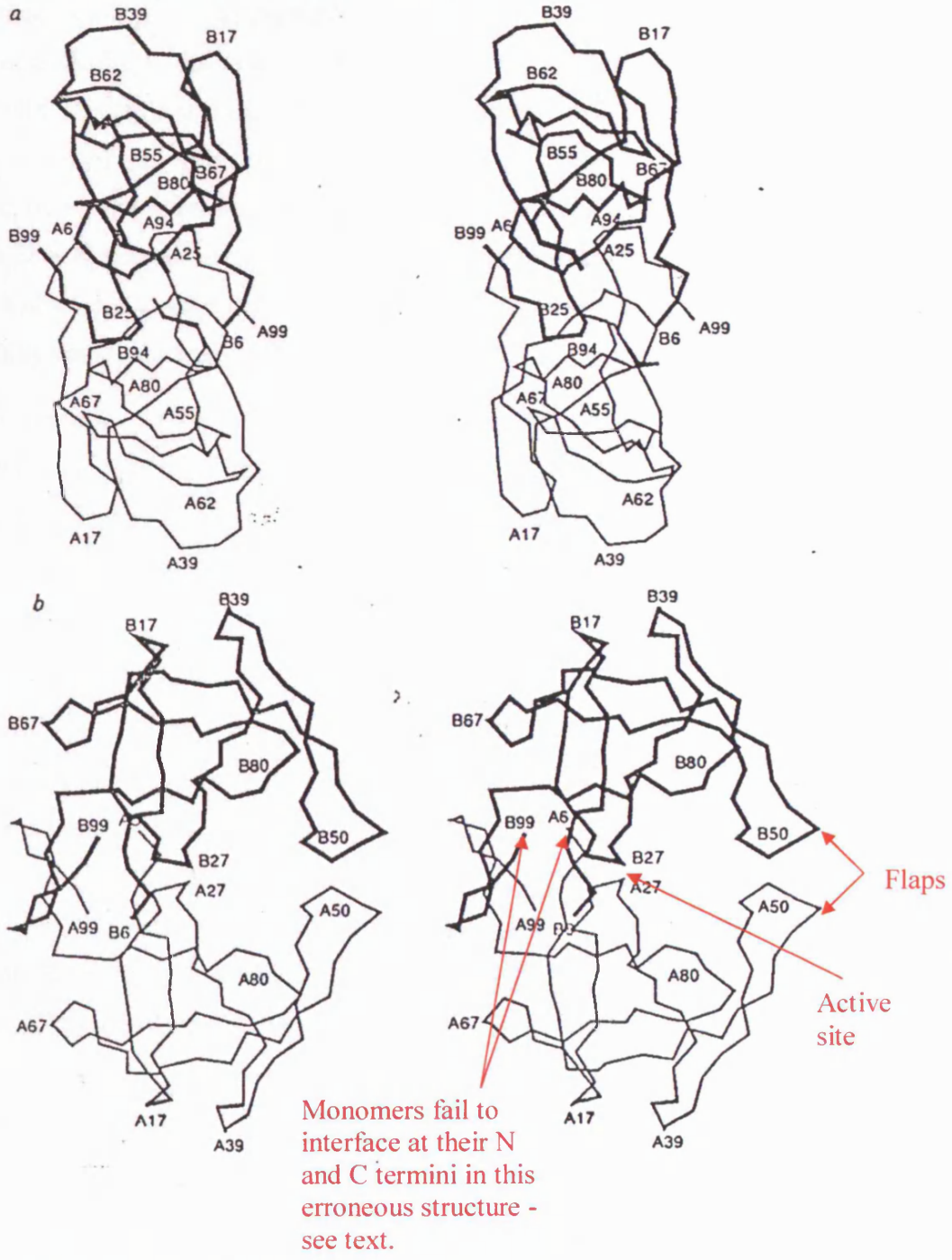


Figure 7 (Navia, Fitzgerald *et al.* 1989) shows a stereo Cα representation of the HIV-1 PR dimer (monomers are arbitrarily labelled A and B, and drawn in thin and thick lines respectively). The top view is along the dimer's twofold axis of symmetry, whilst the bottom view shows a rotation of the first view by 90° about the vertical axis.

Confirmation came when mutation of the conserved Asp 25 residue eliminated protease activity, just as predicted given it to be the catalytic residue of an aspartyl protease (Kohl, Emini *et al.* 1988; Mous, Heimer *et al.* 1988; Seelmeier, Schmidt *et al.* 1988; Darke, Leu *et al.* 1989). More importantly mutagenesis of Asp 25 caused the generation of non-infectious viral particles, demonstrating that maturation and host infection by the retrovirus are dependant on an active PR (Kohl, Emini *et al.* 1988). This clearly established PR as a serious target for antiviral therapy, and confirmed that knowledge gained designing inhibitors of other aspartic proteases could be very usefully employed against HIV PR.

Soon crystals of HIV-1 PR were grown and preliminary data collected (McKeever, Navia *et al.* 1989), and this was swiftly followed by a three dimensional structure for HIV-1 PR solved to 3.0 Å resolution (Navia, Fitzgerald *et al.* 1989; Wlodawer, Miller *et al.* 1989; Lapatto, Blundell *et al.* 1989). It was the first structure of an HIV protein to be solved.

The initial structure by Navia *et al.*, as illustrated in figure 7, was the subject of some controversy following the publication of rival structures by Wlodawer *et al.* and Lapatto *et al.* very shortly afterwards. The cause of this controversy was the fact that an interaction between the monomers at their N and C termini was seen in the later structures (illustrated figs. 8 and 9) but was absent in the earlier one (fig. 7). The main dimerization interface is composed of the interdigitated N- and C-terminal portions of the protease monomers, which together form a four-stranded antiparallel ß-sheet. The later X-ray structures showed that the interface directly involved the N-terminal residues 1-4 and C-terminal residues 96-99 of each monomer, whereas the initial structure by Navia *et al.* lacked residues 1-5 and thus did not reveal the true nature of this interaction. This interface region is highly conserved among HIV-1 isolates and has since become a target for compounds designed to block the assembly of the homodimer or disrupt the dimeric interface, and thus inactivate the enzyme (Boggetto and Reboud-Ravaux 2002; Bowman and Chmielewski 2002).

The structural solutions demonstrated that the PR was a symmetrical homodimer, and that it was similar in structure to the other aspartyl proteases of the pepsin family (Navia, Fitzgerald *et al.* 1989; Wlodawer, Miller *et al.* 1989). It was also similar to other retroviral proteases, such as the Rous Sarcoma Virus (RSV) protease (Miller,

Jaskolski *et al.* 1989). As discussed above, and illustrated in figures 8 and 9, the dimer was stabilised by a four-stranded antiparallel ß-sheet formed by ß-strands at both the N- and C-terminal ends. Also, the active site did indeed form at the interface between the catalytic triads of each subunit, as predicted. A pair of 'flaps' enclosed the active site, each formed by two antiparallel ß-strands connected by a ß-turn at residues 49-52. It was clear from early structural solutions that the position of the flaps differed in the presence and absence of complexed inhibitor (Miller, Schneider *et al.* 1989). In the native enzyme the flaps adopt an open conformation (as seen in fig. 7 b), whereas in the presence of a bound inhibitor they appear to close down (fig. 8) over the active site and substrate binding cleft (Wlodawer and Erickson 1993). Subsequent NMR studies have revealed that the flaps are flexible (Nicholson, Yamazaki *et al.* 1995; Yamazaki, Hinck *et al.* 1996) and therefore able to undergo this conformational change. A water molecule (not to be confused with the nucleophilic water which is bound between the substrate and the catalytic residues) was frequently observed (fig. 10) to be involved in hydrogen bond formation between the flaps and a bound inhibitor (Wlodawer and Erickson 1993). The hydrogen bonding mediated by this water may contribute to distortion of the scissile bond (Pearl 1987; Suguna, Padlan *et al.* 1987) and possibly plays an important role in the mechanism of catalysis (Miller, Schneider *et al.* 1989; Fitzgerald, McKeever *et al.* 1990; Swain, Miller *et al.* 1990; Jaskolski, Tomasselli *et al.* 1991; Harrison and Weber 1994). Specific deletion of these hydrogen bonds resulted in a dramatic decrease in catalytic activity (Baca and Kent 1993). The eukaroytic aspartic proteinases have only a single flap, which interacts with the substrate directly rather than via a water molecule. Consequently it was considered that designing compounds able to displace the structural water molecule might provide the basis for inhibitors specific to the retroviral proteases.

It was eventually shown that careful inhibitor design could achieve displacement of this structural water by a carbonyl group (Lam, Jadhav *et al.* 1994; Lam, Ru *et al.* 1996; Wang, Freedberg *et al.* 1996)[4].

[4] More recently however, evidence suggests that while this structural water plays a role in inhibitor binding, it is not actually involved in the catalytic action of the protease (see Baca and Kent 2000 (Baca and Kent 2000), and references therein).

Figure 8 HIV-1 PR illustration of 3D structure, showing closed flaps, active site with bound inhibitor, and dimerisation domain.

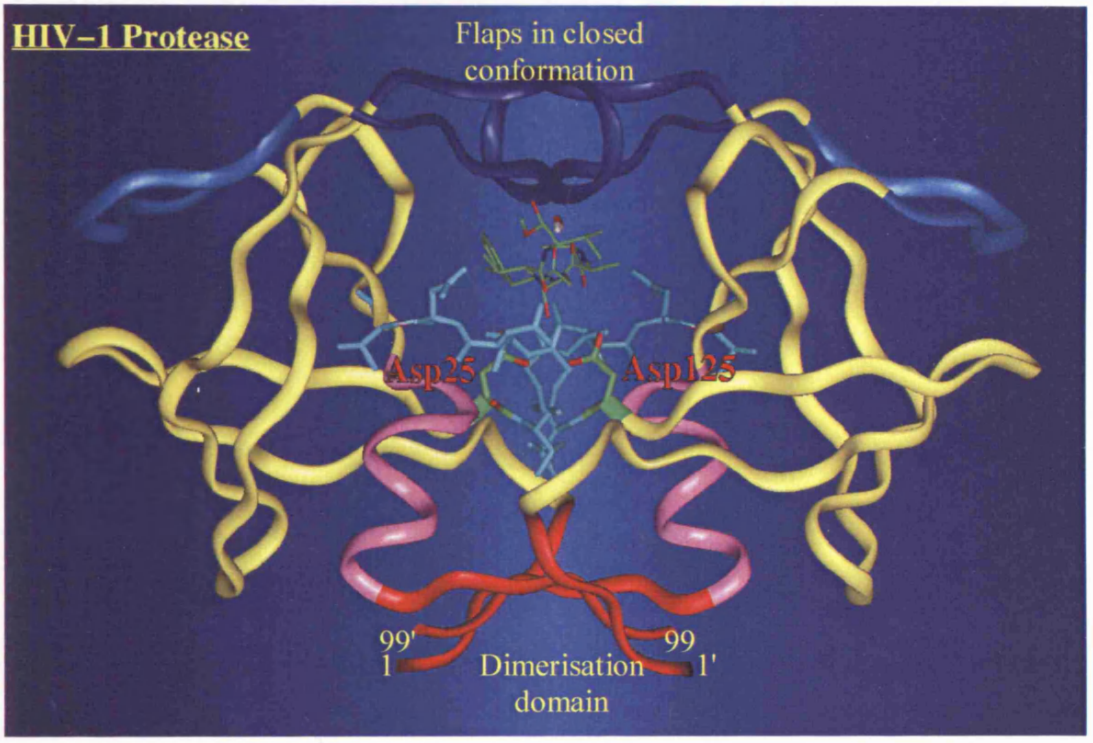


Figure 9 Schematic diagram of dimerisation domain.

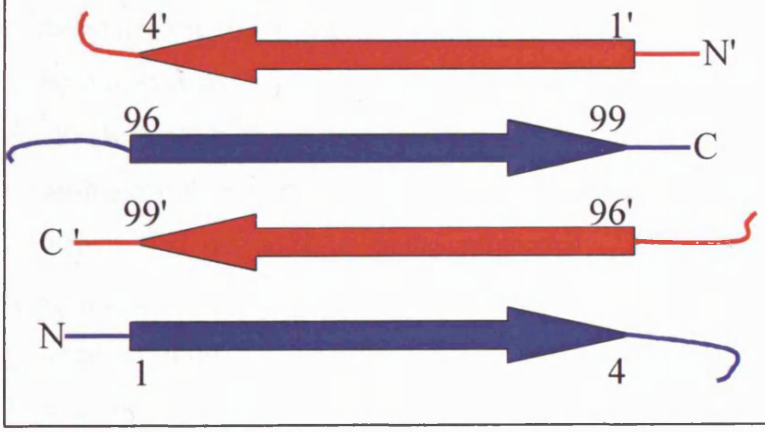


Figure 10 HIV-1 PR active site with inhibitor.

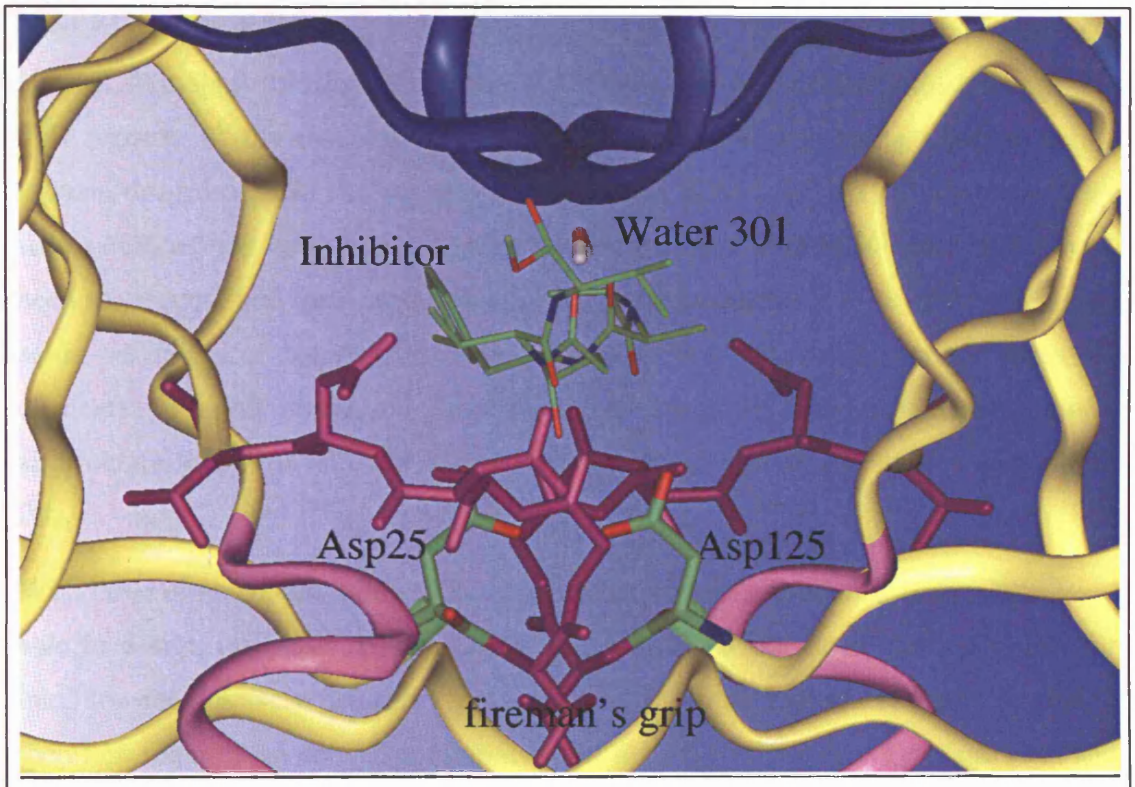


Figure 10 shows a water molecule (water 301) bound between the enzyme flaps and an inhibitor in the protease active site. It may contribute to substrate / inhibitor binding, but probably does not influence catalysis. Asps 25 and 125 are shown in green.

Between 1989 and 1993 over 160 crystal-structures of HIV-1 PR and HIV-1 PR inhibitor complexes were solved (Wlodawer and Erickson 1993). The information derived from these structures was used in the rational development of HIV PR inhibitors (considered later in this chapter, page 85 onwards). A recent query revealed 181 hits from the PDB database when searching "HIV-1 protease" (text search settings: full text and match exact word), and of these 160 were accessible, and 21 were either on hold or being processed (13 of these were dated 2004). Of the structures available 13 were deposited in 2002, 4 in 2003, and 6 in 2004, demonstrating that structural studies of this protein, its mutants and inhibitors are ongoing.

The focus of attention then turned to the structure of HIV-2 PR, since any drugs directed against the proteases should preferably be effective against both serotypes of HIV. A structural solution of the HIV-2 PR was thus one of the initial aims of this PhD project. In the meantime modelling of HIV-2 PR was being used to enable rational drug design in the absence of a structure. HIV-1 and HIV-2 share no more than a 60% overall amino-acid identity. Many isolates of both HIV-1 and HIV-2 have been sequenced, and their proteases have been compared. Only 41% of the PR amino acids are identical between the two serotypes, however, the important structural elements are highly conserved. Since most differences are found in the loop regions, an accurate model of HIV-2 PR could be obtained based on the structure of HIV-1 PR.

With three dimensional models and actual structural data available, researchers were able to design inhibitors to specifically target HIV PR. Obvious starting points for inhibitor designs were pepstatin analogues and substrate analogues. Inhibitor design therefore required an extensive knowledge of the PR recognition / cleavage sites and their interactions with the PR substrate binding cleft. Studies had shown that PR is able to recognise several cleavage sites within the GAG and GAG-POL polyproteins at the junctions between constituent protein subunits (Darke, Nutt *et al.* 1988; Nutt, Brady *et al.* 1988; Schneider and Kent 1988; Darke, Leu *et al.* 1989). Since PR is also part of the POL polyprotein, it must achieve its own release from the POL polyprotein. This autocatalysis, and its inhibition by mutagenesis of the active site, has been demonstrated by the work of Loeb *et al.* (Loeb, Hutchison *et al.* 1989). In the process of extricating itself from the POL polyprotein PR would also free the other viral enzymes - RT and IN. Subsequently the unfettered PR could release MA, CA, NC and p6 from the GAG polyprotein (fig. 11). Cleavage sites within retroviral polyproteins are heterologous, the amino acids surrounding the cleavage sites are different from site to site (fig. 12), and even the dipeptide comprising the site itself (ie. P1-P1′) differs between different sites. Despite this there is a high degree of specificity, for example HIV-1 PR will not cleave RSV cleavage sites (Krausslich, Ingraham *et al.* 1989) despite sharing the same scissile bonds in some cases. Mutation of polyprotein substrates also disrupts cleavage (Hellen, Krausslich *et al.* 1989).

The different scissile site amino acid sequences give rise to varying cleavage efficiency from site to site. This may influence the order in which proteins are

released from their polyprotein. While there is no clear recognition sequence, an approximate consensus can be perceived (fig. 13), and an algorithm for predicting protease sensitive sites has been published (Poorman, Tomasselli *et al.* 1991).

Figure 11 Cleavage site locations in polyproteins.

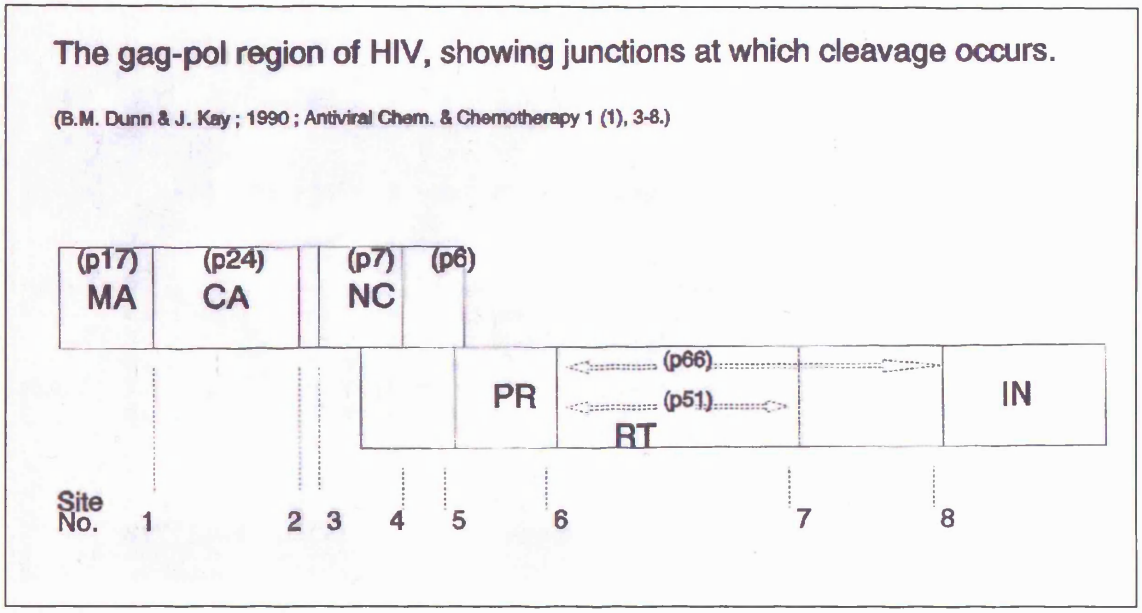


Figure 12 The cleavage site recognition sequences for HIV-1 PR, and predicted sequences for HIV-1 PR and SIV_{AGM} PR.

* = scissile bond. Predictions were made by sequence alignment. For further information relating to retroviral cleavage sites see Dunn *et al.* 1994 (Dunn, Gustchina *et al.* 1994).

HIV-1 known PR cleavage sites (Dunn & Kay 1990).

	-P6 -P5 -P4 -P3 -P2 -P1	*	P'1-P'2-P'3-P'4-P'5-P'6-
1.	-ser-gln-asn-tyr	*	pro-ile-val-gln-
2.	-ala-arg-val-leu	*	ala-glu-ala-met-
3.	-ala-thr-ile-met	*	met-gln-arg-gly-
4.	-pro-gly-asn-phe	*	leu-gln-ser-arg-
5.	-ser-phe-asn-phe	*	pro-gln-ile-thr-
6.	-thr-leu-asn-phe	*	pro-ile-ser-pro-
7.	-ala-glu-thr-phe	*	tyr-val-asp-gly-
8.	-ala-lys-ile-leu	*	phe-leu-asp-gly-

SIV (AGM) possible PR cleavage sites.

	-P6 -P5 -P4 -P3 -P2 -P1	*	P'1-P'2-P'3-P'4-P'5-P'6-	
1.	~850 -Ser-Gln-Asn-Phe	*	Pro-Ala-Gln-Gln-	
2.	~1325 -Gln-Gly-Pro-Lys-Gln-Pro	*	Phe-Lys-Asp-Tyr-Val-Asp-	
3.	~1330 -Val-Asp-Arg-Phe	*	Tyr-Lys-Ala-Ile-	
3.	~1401 -Trp-Met-Thr-Glu-Ser-Leu	*	Leu-Ile-Gln-Asn-Ala-	
4.	~1750 -Gln-Val-Asn-Phe	*	Leu-Gly-Tyr-Gly-	
5.	~1950 -Asp-Arg-Gly-Ile-Phe	*	Phe-Glu-Leu-Pro-Leu-	
6.	~2250 -Val-Pro-Arg-Leu-Val-Met	*	Gly-Gln-Leu-Ser-Gln-Lys-	
7.	~3500 -Phe-Val-Ser-Thr-Pro	*	Pro-Leu-Val-Arg-Leu-	(p51)
8.	~3950 -Ile-Arg-Gln-Val-Leu	*	Phe-Leu-Glu-Lys-Ile-	(p66)

HIV-2 PR possible cleavage sites.

	-P5 -P4 -P3 -P2 -P1	*	P'1-P'2-P'3-P'4-P'5-
1.	~830 -Gly-Gly-Asn-Tyr	*	Pro-Val-Gln-His-
2.	~1410 -Pro-Lys-Glu-Pro	*	Phe-Gln-Ser-Tyr-
3.	~1450 -Tyr-Val-Asp-Arg-Phe	*	Tyr-Lys-Ser-
3.	~1505 -Thr-Gln-Thr-Leu	*	Leu-Val-Gln-Asn-
4.	~1850 -Gln-Ala-Gly-Phe	*	Leu-Gly-Leu-Gly-
5.	~1950 -Asn-Arg-Gly-Leu	*	Ala-Ala-Pro-Gln-
6.	~2380 -Thr-Ala-Leu-Gly	*	Met-Ser-Leu-Asn-
6.	~2385 -Ser-Leu-Asn-Leu	*	Pro-Val-Ala-Lys-
7.	~3500 -Phe-Val-Ser-Thr-Pro	*	Pro-Leu-Val-Arg-Leu-
8.	~3950 -Ile-Arg-Gln-Val-Leu	*	Phe-Leu-Glu-Lys-Ile-

Figure 13 Common characteristics shared by the HIV-1 PR cleavage sites.

PROTEASE CLEAVAGE SITES :- SOME COMMON FACTORS.

Compiled from the data for HIV-1.

I. The consensus sequence :-

(Ser/Thr) - X - Y -Z(Phe/Tyr) * Pro

Where x = small & hydrophobic
y = aromatic or large & hydrophobic
z = small & hydrophobic

* = Scissile site.

II. Seven amino acids (p4 - p'3) are required for specific, efficient cleavage.

III. All eight scissile sites have an overall hydrophobic composition.

A. Residues at positions p1 & p'1 are always hydrophobic.

B. Those at p2 & p'2 are generally hydrophobic or uncharged polar residues.

C. Charged residues are virtually never present at p2/p4 or p'2/p'4.

IV. Differences between the sites cleaved by HIV-1 & HIV-2.

A. HIV-2 generally shows broader substrate specificity.

B. HIV-1 able to cope with aromatic & branched residues at p1 & p'1.

C. HIV-2 less able to cleave substrates with aromatic or branched residues at p1 & p'1, prefers substrates with small substituents at these points.

The active site is responsible for cleavage of peptide bonds, but not for the selection of the bond to be cleaved. The substrate binding sites flanking the active site determine where a substrate is cleaved. These contain hydrophobic pockets that bind the side chains of amino acids adjacent to the substrate cleavage site, and thus determine both substrate selection and rate of turnover. The two flaps (already described) close over the active site primarily to desolvate the bound substrate, but may also contribute to substrate recognition (Miller, Schneider *et al.* 1989; Fitzgerald, McKeever *et al.* 1990; Swain, Miller *et al.* 1990; Jaskolski, Tomasselli *et al.* 1991).

Enzyme kinetics using a variety of assay systems – for example chromogenic peptides such as described by Konvalinka *et al.* (Konvalinka, Strop *et al.* 1990) - began to provide a great deal of information concerning the substrate specificities of the HIV PRs. This was augmented by the development of models and algorithms to predict whether a peptide could be cleaved by the protease or not (Chou 1993; Chou 1993; Chou and Zhang 1993; Chou, Zhang *et al.* 1993; Zhang and Chou 1994; Cai, Yu *et al.* 1998). New substrate analogues were designed with optimised affinity for the substrate binding cleft, and the inclusion of a non-scissile bond at the cleavage site to create a substrate analogue inhibitor. There were technical difficulties involved, however. Firstly, large quantities of pure HIV PR were required for the many assays and structural studies involved in inhibitor testing, and HIV PR is not expressed at high levels in any conventional expression system. This problem was both encountered, and addressed, during the completion of this thesis. Secondly, the peptide substrates and inhibitor compounds were necessarily hydrophobic, and thus extremely insoluble in aqueous solutions. This made them difficult to use in assays, and would also impact upon the bioavailability of any drugs based on them.

Whilst inhibitor development and design will be considered later in this chapter (page 85 onwards), it is useful to consider some aspects of protease inhibition at this stage. Aspartic proteinase activity is unaffected by serine, cysteine or metallo-proteinase inhibitors such as PMSF, divalent cations, and EDTA, but are in general sensitive to a number of inhibitor classes. Diazo compounds such as diazotyl norleucine methyl ester (DAN) react covalently with Asp 215 of porcine pepsin (Rajagopalan, Stein *et al.* 1966; Bayliss, Knowles *et al.* 1969), and this reaction is specific for a functional active site. Epoxides such as 1,2-epoxy-3-(p-nitrophenoxy) propane (EPNP) have been shown to inhibit pepsin (Tang 1971), the inhibitor molecules reacting in a 2:1

mole ratio with the enzyme. A crystal structure of penicillopepsin in complex with EPNP (James, Hsu *et al.* 1977) revealed that two EPNP molecules form ester linkages with the carboxyl groups of the catalytic aspartate residues. Chymosin, renin, and various fungal apsartic proteinases are also inhibited by diazo and epoxy compounds (Morihara and Oka 1981). The archetypal aspartic proteinase inhibitor, however, is pepstatin, a competitive inhibitor having a central statine residue. The isovaleryl-derivative of pepstatin (isovaleryl-Val-Val-Sta-Ala-Sta) is the most commonly used, and indeed is generally referred to as "pepstatin", but it should also be noted that other derivatives - acetyl-pepstatin (Ac-Val-Val-Sta-Ala-Sta) and lactoyl-pepstatin (Lac-Val-Sta-Ala-Sta) – have also been reported (Satoi and Murao 1970; Kay, Afting *et al.* 1982) and shown to inhibit aspartic proteinases (Valler, Kay *et al.* 1985). Although isovaleryl-pepstatin was found to be a relatively poor inhibitor of HIV-1 PR (Seelmeier, Schmidt *et al.* 1988; Giam and Boros 1988; Hansen, Billich *et al.* 1988), and lactoyl-pepstatin even worse (Richards, Roberts *et al.* 1989), the acetyl derivative proved to be a relatively good inhibitor of HIV-1 PR (Richards, Roberts *et al.* 1989). The differences between the derivatives in terms of HIV PR inhibition were subsequently explained by a prediction that the enzyme's P4 subsite was only able to accommodate a small residue or moiety, such as the acetyl (Weber, Miller *et al.* 1989).

It is thought that the tetrahedral hydroxy group of the statine acts as a transition state analogue of the substrate scissile bond carbonyl group (Marciniszyn, Hartsuck *et al.* 1976). This introduces another highly specific type of inhibitor, the transition state analogue inhibitor, the design of which is reliant on knowledge of the enzyme's chemical mechanism.

Elucidation of the aspartic proteinase mechanism.

The mechanism by which the retroviral aspartic proteases achieve catalysis is essentially the same as that employed by the other aspartic proteinases. A complete understanding of the aspartic proteinase catalytic mechanism, and more specifically that of the retroviral aspartic proteases, would facilitate design of effective HIV PR inhibitors. It is generally accepted that the catalysis involves both aspartate residues, and probably also the participation of the water molecule observed bound between the aspartates in many of the PR crystal structures. However, the precise mechanism of aspartic proteinases has been the subject of some controversy.

Early proposals: a covalent enzyme - substrate intermediate.

It was initially suggested that a covalent enzyme-substrate intermediate might be involved. Experiments using pepsin and penicillopepsin were thought to imply a covalent acyl-enzyme intermediate (Takahashi, Wang *et al.* 1974), and the failure of free amino acids to exchange with the enzyme-amino acid intermediate supported the idea that the intermediate was covalent (Wang and Hofmann 1976). Various workers tried to demonstrate the presence of a covalent intermediate, predominantly using porcine pepsin and penicillopepsin. There were attempts to trap a covalent intermediate (Cornish-Bowden, Greenwell *et al.* 1969), stopped-flow experiments (Hofmann and Hodges 1982), cryoenzymological studies (Dunn and Fink 1984; Hofmann, Fink *et al.* 1984), and Isotope incorporation experiments using $H_2{}^{18}O$ (Antonov, Ginodman *et al.* 1978; Antonov, Ginodman *et al.* 1981), but they all failed to provide any support for the idea. The isotope incorporation experiments, on the contrary, showed that a general base catalysed attack by H_2O was more likely (Antonov, Ginodman *et al.* 1981). Crystallographic structures of rhizopuspepsin and penicillopepsin finally demonstrated that the formation of a covalent acyl- or amoni-enzyme intermediate is precluded by steric hindrance (Bott, Subramanian *et al.* 1982; James, Sielecki *et al.* 1982). However, crystal structures of aspartic proteinases consistently revealed the presence of a water molecule associated with the catalytic aspartates.

Early proposals – general base assisted nucleophilic attack by water.

The observation of an active site water molecule supported the proposal that catalysis proceeds via a general base assisted attack by a water molecule and does not involve a covalent intermediate (Hofmann, Fink *et al.* 1984). It was thought possible that the observed water molecule could be the required nucleophile.

Early crystallographic studies of aspartic proteinases (Hsu, Delbaere *et al.* 1977; Jenkins, Tickle *et al.* 1977; Subramanian, Liu *et al.* 1977; Bott, Subramanian *et al.* 1982; Blundell, Sibanda *et al.* 1983; James and Sielecki 1983; James and Sielecki 1986) suggested that one of the catalytic aspartate residues - Asp 32 - was ionised (residues are numbered according to the system applied to Porcine Pepsin by Tang *et al.* PNAS 1973) while the other - Asp 215 - was protonated, and contemporary explanations of the potential mechanism reflected this (James, Hsu *et al.* 1977; Blundell, Jones *et al.* 1980). Among these, the charge relay mechanism of James *et al.* (James, Hsu *et al.* 1977) was later discounted because the large conformational changes by the enzyme that it required were shown to be unlikely (Pearl and Blundell 1984).

Higher resolution studies of penicillopepsin (James and Sielecki 1983) and endothiapepsin (Pearl and Blundell 1984) suggested that the aspartates were in fact equivalent and symmetrical about a local 2-fold axis. The higher resolution data showed the active site water molecule to be within hydrogen bonding distance of all four of the catalytic carboxyl oxygens (James and Sielecki 1983; Pearl and Blundell 1984). Also revealed was the close proximity of the carboxyl groups of the catalytic aspartates to one another, leading to the assumption that a proton must lie between these negatively charged groups since they would otherwise be expected to repel each other (Pearl and Blundell 1984). These observations and assumptions depicted a symmetrical image of the native active site, and it was suggested that this may indicate that an average of possible arrangements (fig. 14) was being observed (Pearl and Blundell 1984).

Figure 14. Possible arrangements of active site water and proton.

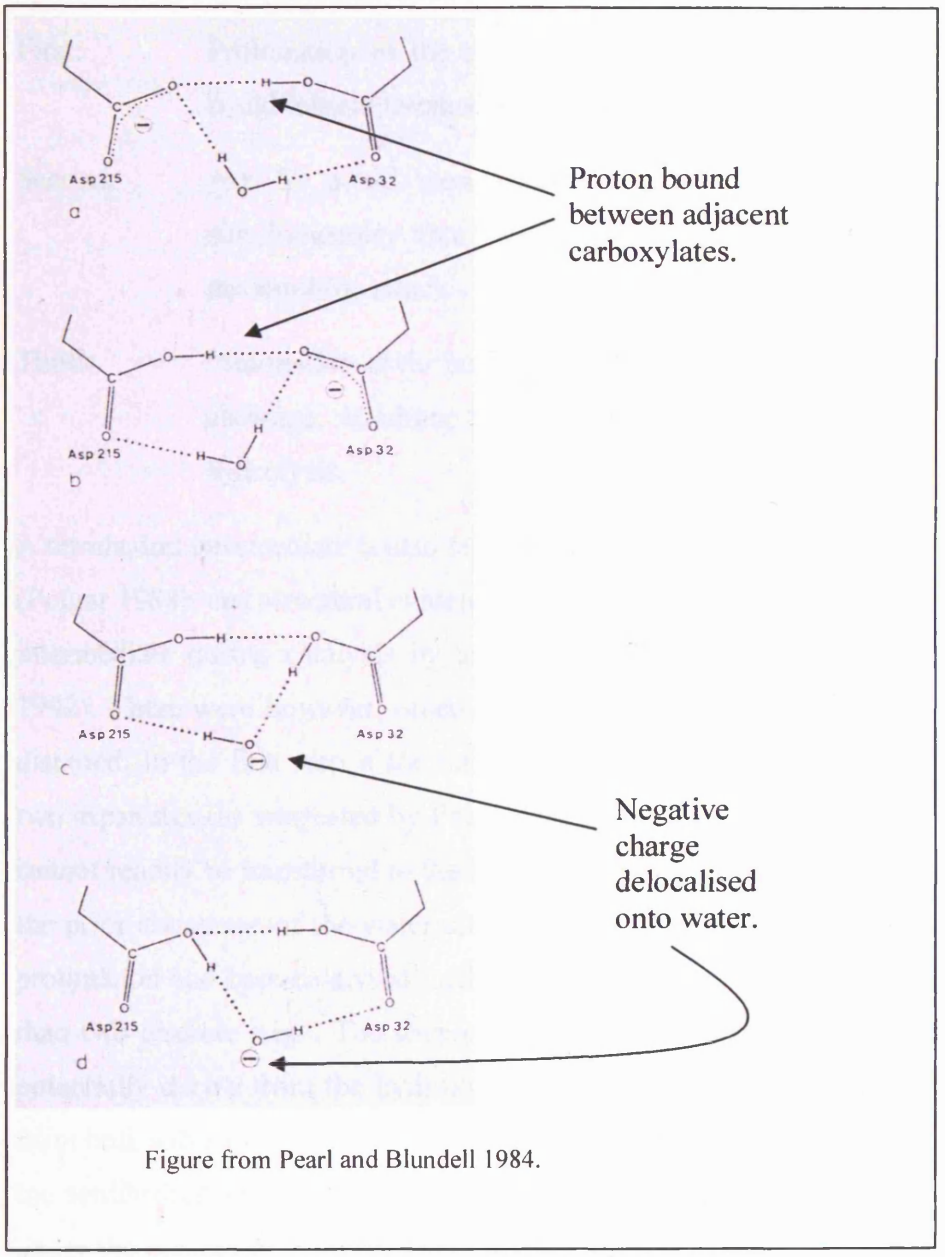


Figure from Pearl and Blundell 1984.

In this case the proton remains bound between the adjacent carboxylates, whilst the negative charge from the adjacent residues is delocalised onto the water. This would indicate that the water could potentially be a strong nucleophile, and thus initiate peptide bond cleavage.

The mechanisms put foward by Pearl and Blundell, James and Sielecki (James and Sielecki 1983; Pearl and Blundell 1984) were based on the structure of penicillopepsin in complex with pepstatin (James and Sielecki 1983) and incorporated this potentially nucleophilic water molecule.

The proposals incorporate the following generalised mechanistic steps:

First: Protonation of the substrate carbonyl oxygen, probably by Asp 215 could initiate tetrahedral intermediate formation.

Second: Asp 32 could then abstract a proton from the bound water, and simultaneously that water could engage the carbonyl carbon in a nucleophilic attack - forming a tetrahedral gem-diol intermediate.

Third: Protonation of the peptide nitrogen atom would culminate in C-N bond cleavage, resulting in the amine and carboxyl products of peptide hydrolysis.

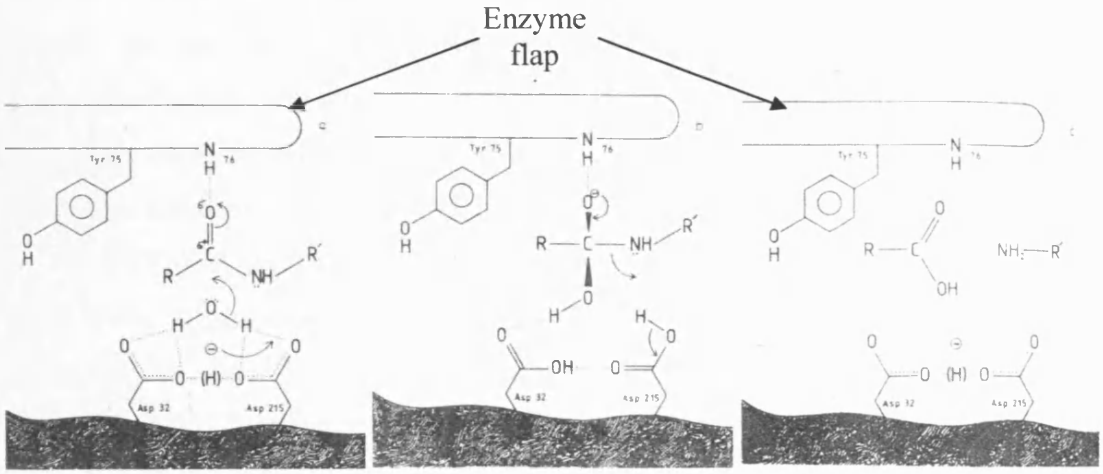
A tetrahedral intermediate is also featured in the mechanisms of other protease classes (Polgar 1988), and structural evidence supports the existence of a gem-diol tetrahedral intermediate during catalysis by aspartic proteinases (Veerapandian, Cooper *et al.* 1992). There were however, other aspects of the proposed mechanism that could be disputed. In the first step if the enzyme associated proton is H-bonded between the two aspartates (as suggested by Pearl and Blundell) and held by a negative charge, it cannot readily be transferred to the less basic neutral peptide carbonyl oxygen without the prior assistance of the water attack (Polgar 1987). It must be assumed then, that protonation and base-catalysed nucleophillic attack are simultaneous processes rather than two discrete steps. The source of the proton in step 3 is also uncertain. It may potentially derive from the hydroxyl of the tetrahedral intermediate via Asp 215, or from bulk solvent. However, Polgar argued that protonation of the leaving nitrogen of the tetrahedral intermediate from bulk solvent (specific acid catalysis) is unlikely under the conditions favoured by aspartic proteinases, so that this is eliminated as a source of the proton in step 3 (Polgar 1987). Alternatively the proton abstracted from the attacking water could be passed between the aspartates to protonate the leaving nitrogen of the tetrahedral intermediate (Polgar 1987). Given the high symmetry around the two active site carboxyl groups (Pearl and Blundell 1984) it may be that the dyad proton is able to bind either carboxyl in the free enzyme. The mechanism would then involve two concurrent and symmetrical proton transfers for both the formation and breakdown of the tetrahedral intermediate, or 'push-pull' catalysis (Polgar 1987).

The scissile C-N bond of the peptide substrate may be distorted by the interaction of substrate residues elsewhere binding the active site cleft and enzyme flap residues, a phenomenon known as substrate strain (Pearl 1985; Davies 1990). It had previously been suggested that the strain imposed by substrate binding to pepsin could contribute to peptide bond cleavage (Fruton 1976). The lone pair on the peptide nitrogen is ordinarily delocalised across the C-N bond providing it with characteristics akin to a double bond. On binding of substrate to the enzyme cleft, interaction of substrate residues remote from the scissile bond with their adjacent cleft residues may cause distortion of the scissile bond. This may have the effect of reducing nitrogen lone pair delocalisation so that reversion to a state more characteristic of a single bond would occur. This change in the C-N bond would render it more susceptible to cleavage by nucleophillic attack (Pearl 1987).

Subsequently enzyme inhibitor complexes have been interpreted in terms of substrate strain (Suguna, Padlan *et al.* 1987; Jaskolski, Tomasselli *et al.* 1991). However, crystallographic studies of HIV-1 and SIV proteases complexed with both substrate and the hydrolysis products of that substrate have allowed modeling of the substrate enzyme complex (Rose, Craik *et al.* 1996). The resultant models lack any strain in the scissile peptide bond, and indicate the bond would not actually be deformed on substrate binding (Rose, Craik *et al.* 1996).

Figure 15 The catalytic mechanism of aspartic proteinases as proposed by L.H. Pearl 1987.

Adapted from figures within: Pearl L.H., 1987, FEBS Letters, Vol. 214, no.1, pp. 8-12



1. Nucleophilic attack by water on carbonyl carbon at scissile bond of peptide.

2. Tetrahedral transition state.

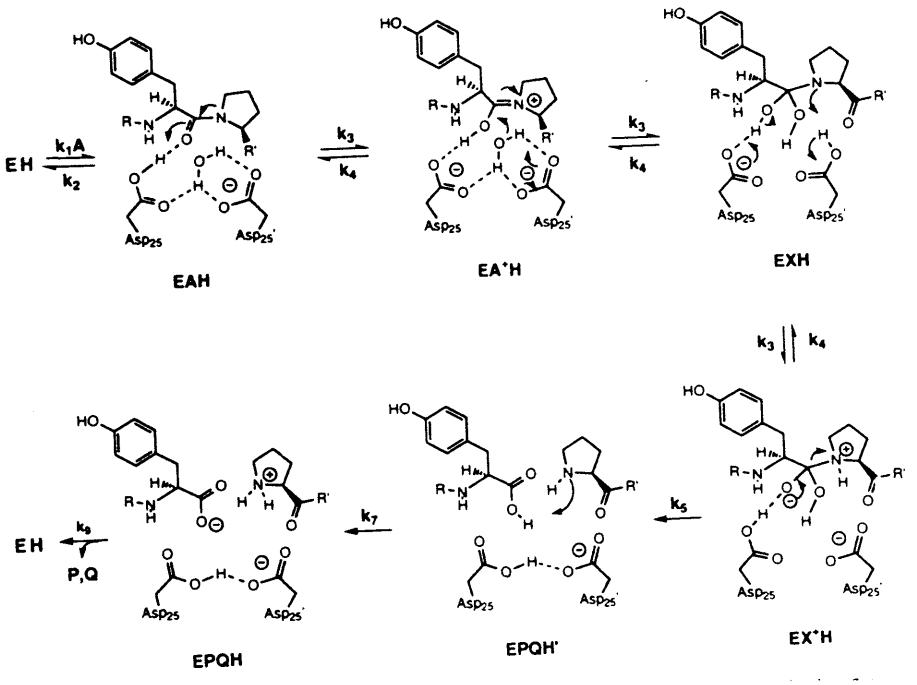
3. The transition state collapses to form the products of hydrolysis.

This figure illustrates the steps in a proposed mechanism for aspartic proteinases, based on penicillopepsin, which incorporated the possibility that the scissile C-N bond is put under strain as the substrate interacts with remote substrate cleft and flap residues.

Mechanisms with particular relevance to the retroviral proteases are proposed.

As structural and kinetic data for the retroviral proteases and their complexes with inhibitors became more prevalent, consideration of the mechanism with specific respect to retroviral proteases became possible. Detailed potential mechanisms, specific for the HIV protease, have been described taking into account all the available structural and kinetic data available at the time (Lin, Lin *et al.* 1994; Meek, Rodriguez *et al.* 1994). Each of these involves the nucleophilic attack by the bound water and the formation of a tetrahedral intermediate. They are also both referred to by the BRENDA protease database (a "comprehensive enzyme information system" to be found at http://brenda.bc.uni.koeln.de) as the references of choice when wishing to understand the HIV PR mechanism. Consequently they have been included here (figs. 16 and 17) as alternate mechanism proposals contemporary with the work described in this thesis, and as the basis for most further developments in attempting to establish the absolute mechanism.

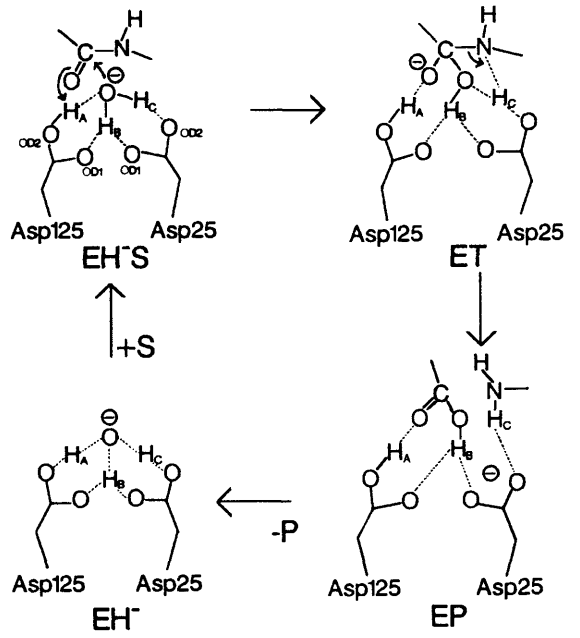
Figure 16 Catalytic mechanism of HIV-1 PR proposed by Meek *et al.* 1994.



Proposed chemical mechanism for HIV-1 protease-catalysed peptidolysis of:
Ac-Ser-Gln-Asn-Tyr-Pro-Val-Val-NH_2
based on crystallographic data, previous kinetic results, and the present isotopes effects data.

Figure taken from Meek T.D., Rodriguez E.J., and Angeles T.S., 1994, Methods in Enzymology, Vol. 241, pp127-156

Figure 17 Catalytic mechanism of HIV-1 PR proposed by Lin *et al.* 1994.



Proposed mechanism of hydrolysis of HIV-1 protease.

In the free enzyme, EH^-, two active site aspartyl groups (Asp-25 and Asp-125) are assumed to share a proton and a water molecule. The negative charge is assigned to the water oxygen. The enzyme-substrate complex is presented as EH^-S, three protons in the active site are marked H_A, H_B, and H_C respectively. The transition state and the enzyme-product complex are marked ET and EP respectively.

Figure taken from Lin X-L, Lin Z-Z, and Tang J, 1994,
Methods in Enzymology, Vol. 241, pp195-224

The mechanism described by Lin *et al.* is essentially that of the eukaryotic aspartic proteinases as described by Davies in 1990 (Davies 1990). It differs in that the retroviral enzyme is a dimer of identical subunits (whereas the eukaryotic enzymes are monomeric and asymetrical), and this confers upon the native active site a symmetry of both structure and charge distribution which is absent in the eukaryotes. However, the initial state (EH) adopted by Lin *et al.* is actually reminiscent of the symmetry proposed by Pearl (Pearl 1987) when describing the general mechanism of aspartic proteinases based on work involving the microbial enzyme endothiapepsin (fig. 15). Essentially both could be describing a situation in which a hydronium (H_3O^+) ion is effectively shared between the negatively charged carboxyls of the catalytic aspartates. In the process the charge from the carboxyls becomes transferred to the oxygen of the hydronium ion. On binding of the substrate this symmetry is lost in both mechanisms (figs. 15 and 17).

Investigation of the proton transfers involved in catalysis.

The investigations thus far considered have established the role of a water molecule in the attack and hydration of the peptide bond in addition to the formation of a tetrahedral intermediate. The consensus of the many mechanisms proposed is that reaction procedes by means of acid-base catalysis, in which a water molecule disrupts the peptide bond hybridisation by engaging in a nucleophilic attack on the substrate carbonyl carbon. This nucleophilic attack results in the formation of a gem-diol tetrahedral intermediate in place of the carbonyl group, and the rearrangement of the amide nitrogen hybridisation from sp^2 to sp^3, i.e. the nitrogen lone pair ceases to be delocalised across the peptide bond. The next stage involves protonation of the nitrogen, and proton release from the gem-diol, with resultant break down of the peptide bond. However, precisely how the proton exchanges implicit in these steps procede is unclear. The possible permutations in terms of proton transfer have lead to a plethora of suggestions with respect to this aspect of the mechanism, and they are yet to be defined absolutely.

Crystallographic and molecular dynamic studies have since focused on this problem. Since the rational design of inhibitors to maximise their interaction with the catalytic aspartate carboxyls requires accurate information about the ionisation states and pKa values of these carboxyl groups, both in the free enzyme and during the catalytic process.

Theoretically the catalytic Asp dyad of the free enzyme can adopt one of three protonation states, these being deprotonated, monoprotonated or diprotonated. In addition any of these states could be adopted as catalysis progresses. Which of these states is prevalent at any given point in the reaction must be determined to fully understand the proton exchanges inherent in the mechanism. However, since protons cannot be visualised via the electron density maps employed to determine protein structures following X-ray crystallographic analysis, this information has proved difficult to obtain. Consequently the protonation states of the aspartate carboxyls have been inferred from other available structural and biochemical information, such as the proximity of the carboxyls to each, rather than from direct observations. Furthermore, molecular dynamic studies have predicted varying aspartyl protonation states dependant on the external pH and on interactions between the carboxyl groups and any bound compound (Harte and Beveridge 1994). This has meant that questions

concerning the nature of proton transfers occuring during catalysis, and thus an understanding of the precise mechanism, have not been resolved by the numerous crystallographic studies. It has become possible to elucidate protonation states by means of NMR spectroscopy. However, NMR measurements are made in solution, consequently the dynamic nature of proton transfer makes the elucidation of an enzyme's protonation state at any given point in the hydrolysis process extremely difficult. Nevertheless, NMR studies of HIV-1 PR free enzyme and inhibitor complexes have shown that various protonation states can exist, supporting the molecular dynamic predictions. NMR studies of an HIV-1 PR complex with a two fold symmetric inhibitor demonstrated that Asp25 and Asp25' (aka Asp125) carboxyls are both protonated over the pH range 2.2 - 7.0 (Yamazaki, Nicholson *et al.* 1994). Alternatively a separate NMR study on a complex with an asymmetric inhibitor showed that Asp25 was protonated, whereas Asp25' (Asp125) was not (Wang, Freedberg *et al.* 1996). These studies support the notion that the prevailing catalytic carboxyl protonation states are dependent upon the specific interactions involved between them and a given inhibitor. Ultimately, neutron diffraction studies are better suited to the investigation of enzyme protonation states, consequently this technique and its application to the study of aspartic proteinases is considered later in this chapter (See Future mechanistic investigations; Neutron diffraction, page 81).

Figure 18 Illustration of the catalytic carboxyl protonation states considered in *ab initio* dynamic studies by Piana *et al.*

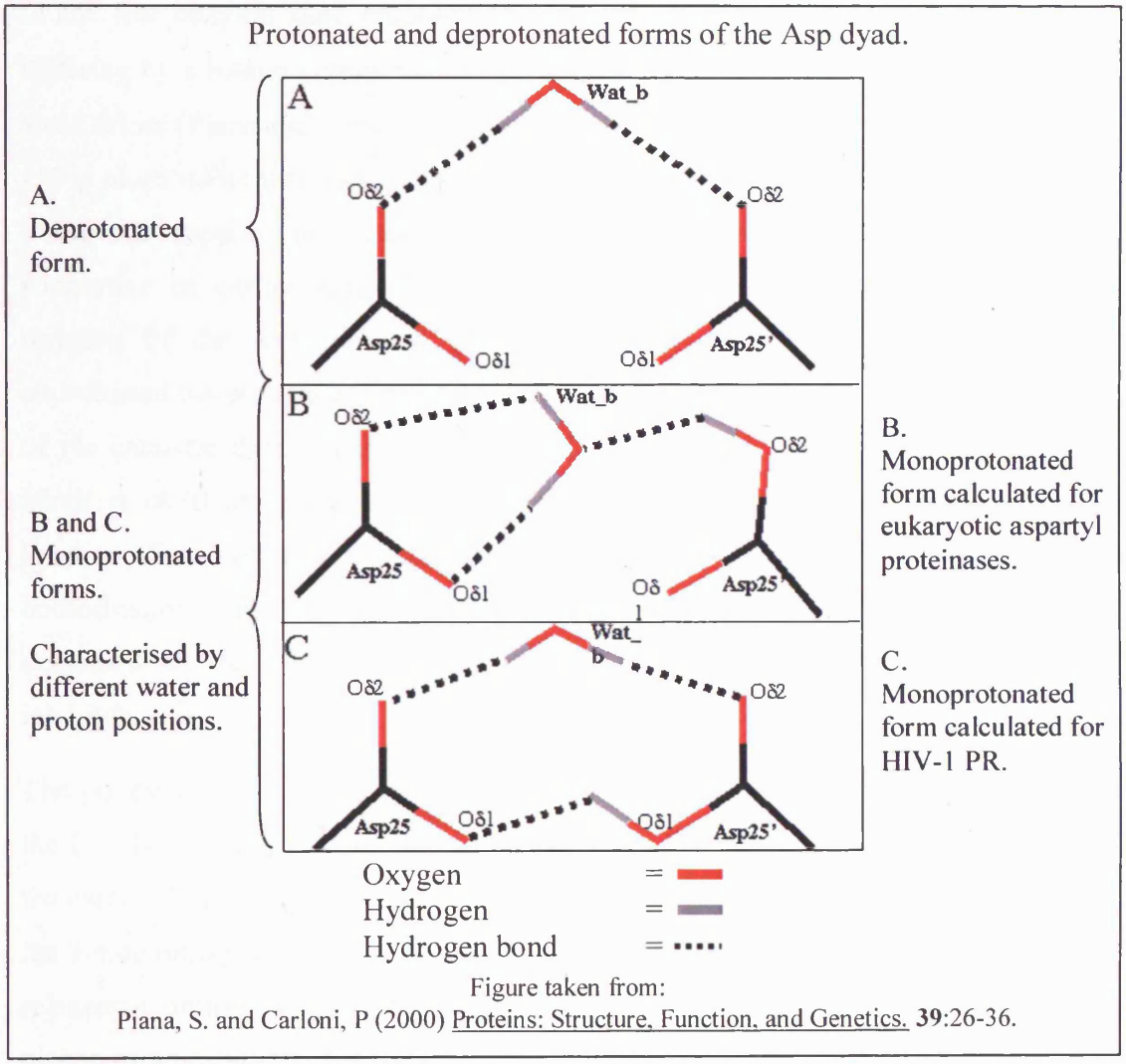


Figure taken from:
Piana, S. and Carloni, P (2000) Proteins: Structure, Function, and Genetics. **39**:26-36.

The deprotonated free enzyme state has already been considered by Pearl and Blundell who concluded that the close proximity of the catalytic carboxyls meant that they would be expected to repel each other unless a proton lay between them (Pearl and Blundell 1984). This supposition has recently been supported by an *ab initio* dynamic study (Piana and Carloni 2000) of the free HIV-1 protease which concluded that in the unprotonated state (fig. 18 Conformation A) the aspartyl groups would repel each other to such an extent that they would be forced to adopt a highly distorted conformation which would be very unstable, and has not been observed experimentally. The same study also investigated the possible active site conformations that may be adopted by the free HIV-1 PR in the monoprotonated state. These are characterised by their differing water and proton positions (fig. 18

Conformations B and C). Conformation B has been shown to prevail in the free eukaryotic aspartic proteinases (Beveridge and Heywood 1993). It is also reminiscent of the free enzyme state proposed by Meek *et al.* (Meek, Rodriguez *et al.* 1994), differing by a rotation about the Cβ-Cγ bond of the protonated Asp. However, Piana and Carloni (Piana and Carloni 2000) concluded that for HIV PR conformation C (fig. 18) is more stable than conformation B - the reverse of the situation prevailing in the eukaryotic aspartic proteinases (Beveridge and Heywood 1993). This result is supportive of earlier suggestions that a proton is located between the adjacent oxygens of the Asp carboxyl groups (Pearl and Blundell 1984), and that the coordinated water molecule makes hydrogen bonds with both the carboxyl side chains of the catalytic dyad (Pearl 1987) rather than with only one as later suggested by Meek *et al.* It also concurs with the crystallographic observations concerning the location of the coordinated water and proximity of the carboxyl groups. The various conformations observed by NMR studies of enzyme inhibitor complexes can be explained by the influence on the free-state exerted by the binding of a given inhibitor.

The puzzle remaining concerns the precise means by which the leaving nitrogen of the C-N bond is protonated, and the peptide bond broken. The nucleophilic attack on the carbonyl carbon results in gem-diol tetrahedral intermediate formation, and causes the amide nitrogen to change its hybridisation from sp^2 to sp^3. The diol subsequently releases a proton, whilst the nitrogen gains one. The diol most likely donates its proton to an aspartate, but the nitrogen could receive a proton either from an active site carboxyl group or from a nearby water molecule (Veerapandian, Cooper *et al.* 1992). Difluoroketone inhibitors (-CO-CF_2-) readily become hydrated (-$C(OH)_2$-CF_2) and thus mimic the hydroxyls of a gem-diol intermediate. The structure of HIV-1 protease in complex with an inhibitor containing a central difluoroketone in a gem-diol form has been used to mimic a substrate peptide bond following its hybridisation collapse, in order to ascertain the means by which the leaving nitrogen is protonated (Silva, R.E. *et al.* 1995; Silva, Cachau *et al.* 1996).

A two step substrate binding process (fig. 19) has been suggested which would involve conformational change on the part of the substrate itself (Silva, Cachau *et al.* 1996), in the mode described by Furfine *et al.* (Furfine, D'Souza *et al.* 1992). In this scenario only half the substrate binds optimally to the enzyme cleft in the first instance. The reaction procedes by means of nucleophilic attack on the carbonyl carbon from the bound water (fig. 19 I), a gem-diol tetrahedral intermediate is formed, (fig. 19 II) and the peptide bond is hydrolysed - just as predicted by previous mechanistic proposals (fig. 19 IV). In the process a proton translocation occurs, so that the previously charged aspartate becomes protonated and vice versa (fig. 19 III). Additionally the changes wrought on the hybridisation state of the amide nitrogen allow rotation about the C-N bond (fig. 19 II). As a result the substrate adopts a second conformation in which the remainder of its residues can now bind the enzyme cleft optimally (fig. 19 III). The amide nitrogen's lone pair is consequently presented towards the newly protonated aspartate carboxyl, and is thus able to accept the proton (fig. 19 III). Simultaneously one of the gem-diol hydroxyls donates a proton (fig. 19 III) to the recently charged aspartate (towards which both the hydroxyls of the diol are orientated). The proton exchange described results in rupture of the peptide bond and regeneration of the initial protonation state of the enzyme (fig. 19 IV and I).

Figure 19 Proposed mechanism involving two-step binding process and conformational change by the substrate.

Catalytic mechanism of HIV-1 Protease

as proposed by Silva *et al* 1996

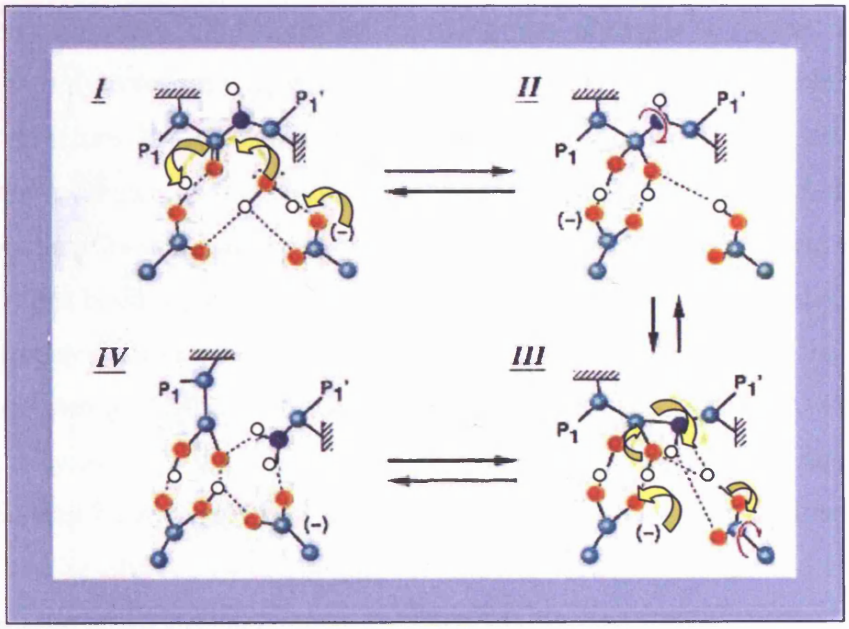


Figure taken from :

Silva A.M., Cachau R.E., Sham H.L, Erickson J.W., 1996, J. Mol. Biol., 255, pp. 321-346

Key:

Arrows; Red arrows indicate rotation, yellow arrows indicate electron transfer.

Spheres; Oxygen is red, Hydrogen is white, Carbon is green, Nitrogen is dark blue.

I Nucleophilic attack by water.

II Tetrahedral intermediate formed, rotation occurs about C-N bond.

III Substrate adopts second conformation, amide lone pair accepts a proton, proton exchange occurs between gem-diol hydroxyl and charged aspartate.

IV Peptide bond broken, initial enzyme protonation state regenerated.

Structural studies have shown that protonation of the leaving nitrogen by the nearest aspartate is procluded by the likely distance (4 Å) between them (Rose, Craik *et al.* 1996). Protonation of the nitrogen from bulk solvent has also been rejected as unlikely under the pH conditions favoured by the enzyme (Polgar 1987). A conformational change in the substrate via the two step binding mechanism proposed by Silva *et al.* would overcome these objections by allowing the nitrogen to come into close proximity with a protonated aspartate. It would therefore seem a plausible proposal. However, structure based modeling of the substrate enzyme complex has not produced any evidence of this conformational change (Rose, Craik *et al.* 1996). On substrate binding the same study suggests that the substrate carbonyl enters a position within hydrogen bonding distance of the protonated apartate, and causes displacement of the coordinated water which adopts a position in close proximity to the carbonyl carbon, and hydrogen bonded to the other aspartate (Rose, Craik *et al.* 1996). The close proximity of the water to the carbonyl carbon at this early stage suggests that nucleophilic attack may commence immediately the substrate is bound (Rose, Craik *et al.* 1996). The results therefore support the notion of Polgar that proton abstraction and nucleophilic attack probably occur concurrently (Polgar 1987).

Overview of the aspartic proteinase mechanism.

The various proposals discussed above are here distilled into an overall view of the most likely mechanism. It is probable that the free enzyme active site exists in a symmetrical state with the pair of catalytic aspartates adopting a monoprotonated state with the proton deployed between the two carboxyls (bound to one, and hydrogen bonded to the other - or vice versa) acting as a buffer between the two repellant groups. A water molecule is symmetrically coordinated by hydrogen bonds to both aspartates. On approach and binding of a substrate, the water is displaced and rotations about the Cβ-Cα bonds of the aspartates are induced, in conjunction with a series of proton tranfers, brought about by the proximity of the δ-negative substrate carbonyl oxygen. The subsequent rearrangement to form the Michaelis complex allows the water to mount a nucleophilic attack on the delta positive carbonyl carbon. In the process a gem-diol tetrahedral intermediate is formed as a proton is transferred to the substrate carbonyl oxygen from the nearest aspartate, and a proton is abstracted from the water by the opposing aspartate, nearest the peptide bond. Rotation about the Cβ-Cα bond of the asapartate abstracting the proton allows this proton to be positioned between the opposing catalytic residues, but at the same time shifts the unprotonated oxygen of that carboxyl into proximity with the tetrahedral intermediate. The peptide nitrogen is protonated by one of the gem-diol hydroxyls, and the peptide bond breaks, releasing the products. The active site reverts to its free state, a new coordinated water molecule having been provided from the bulk solvent. The symmetry of the active site allows either aspartate to be protonated initially, and thus allows an asymmetrical substrate to be processed in either of two orientations - provided there is no conflict in terms of interaction with the substrate binding pockets. This explains why protonation states have been seen to vary according to the inhibitor complex studied, and the efficacy of symmetrical inhibitors.

Future mechanistic investigations.

The quest for an absolute answer to the mechanism question continues, using super computers and techniques such as *ab initio* Hartree-Fock theory to calculate reaction profiles for small, critical, portions of the enzyme substrate complex. As discussed, the major issues remaining to be elucidated concern aspects of protonation and proton transfer. High resolution X-ray diffraction may provide a solution, but a technique that is ideally suited to the task is Neutron Diffraction.

Neutron Diffraction

Neutron diffraction is the phenomenon associated with the interference processes which occur when neutrons are scattered by the atomic nuclei within a sample – just as X-ray diffraction involves the scattering of X-rays by atomic electrons. Electrons, Neutrons and X-rays each have distinct properties that suit each to specific purposes. Their interaction with matter decreases in the order Electrons > X-rays > Neutrons. Most electron diffraction is performed with high energy electrons whose wavelengths are orders of magnitude smaller than the inter-planar spacing in most crystals. Electrons are charged, light particles and their penetration into solids is very limited. Since electrons exhibit a strong interaction with matter electron diffraction is essentially a surface technique. In contrast, X-ray and Neutron diffraction techniques are well suited to structural investigations. Neutrons are neutral particles each with spin 1/2, a magnetic moment of 1.9132 nuclear magnetons, and a half-life of approximately 888 s. Having a magnetic moment causes them to be sensitive to magnetic ordering within a solid.

Relative to X-ray diffraction, Neutron diffraction is a fairly new technique - the first neutron diffraction analysis of a protein structure was completed by Schoenborn in 1969 on myoglobin (Schoenborn 1969), and it is still quite uncommon. This is largely because it requires the generation of high thermal-neutron fluxes (a thermal neutron is defined as a neutron possessing a kinetic energy of about 0.025 eV). These can only be obtained from steady state nuclear reactors or from particle accelerators called spallation sources. Further details concerning source of high thermal-neutron fluxes, and background information on neutron diffraction can be found in the Appendix (Appendix: General Introduction - supplementary information; Neutron Diffraction, page 334).

Although X-ray diffraction has certain advantages over neutron diffraction - one being the much higher intensity of the sources available (the greater intensity of X-ray sources allows the use of smaller crystals and facilitates solution of the phase problem) - the principle distinction between X-ray and Neutron diffraction involves the location of hydrogen atoms. Since X-rays are scattered by the atomic electron cloud, atoms with fewer electrons scatter X-rays less effectively. Hydrogen atoms can therefore only be detected by high precision experiments – impossible in the case of macromolecular crystallography. It is also difficult to distinguish neighbouring atoms of the same row of the periodic table and impossible to distinguish between different isotopes of the same element. Hydrogen represents half of all the atoms in a typical protein, yet a survey of the PDB revealed that only one percent of proteins crystallise well enough to provide sufficient resolution to detect hydrogen atoms using synchrotron X-ray beams.

The greatest advantage of neutron diffraction over X-ray diffraction is its ability to establish the exact position of hydrogen atoms. Since neutron beams interact more strongly with nuclei than do x-rays, neutron diffraction is more useful than x-ray diffraction for determining proton positions. As a result, neutron diffraction can provide data concerning the hydrogen bonds or protonation states of histidine and acidic residues – as demonstrated with lysozyme (Mason, Bentley and McIntyre 1984), the conformations of methyl groups, the solvent structure of the protein, and the degree of protection of particular hydrogen atoms within the structure.

Neutron diffraction is particularly useful when applied to the elucidation of enzyme mechanisms. The best currently accepted mechanisms are largely based on X-ray images of inhibitor structures, but the active-site hydrogen atoms cannot be definitively located by current X-ray analyses, leaving an incomplete picture of enzyme catalysis. For example, Myoglobin was studied by neutron diffraction to compare similarities in the modes of CO (Hanson and Schoenborn 1981) and O_2 (Phillips and Schoenborn 1981) binding. The imidazole of His64 (E7) was only found to be protonated when O_2 was bound. However, Trypsin provides the most prominent example of this application to date. The mechanism of the catalytic triad of serine proteases was finally solved by a neutron diffraction experiment on Mono-Iso-propyl-Phosphate (MIP) inhibited trypsin, which mimics the real enzyme-substrate intermediate (Kossiakoff and Spencer 1981). Kossiakoff and Spencer performed a

neutron structure analysis at a resolution of 2.2 Å on bovine trypsin covalently inhibited by the transition-state analogue MIP. The unique ability of neutron diffraction to locate hydrogen atoms allowed the protonation states of the catalytic site residues (Asp102 and His57) to be determined. It was demonstrated that the proton of the catalytic site resided on the imidazole of His57 and not on the carboxylate of Asp102. Since the bound MIP group mimicked the tetrahedral intermediate structure, the observed protonation states represented those prevailing at the most crucial stage in hydrolysis. This resolved a much debated mechanistic issue by showing conclusively that the catalytic base in the transition state of the reaction is His57, not Asp102, improving understanding of the serine protease hydrolysis mechanism. The findings were later confirmed by ^{15}N labelled NMR studies (Bachovchin 1985). Kossiakoff and Spencer also completed a detailed examination of the stereochemical interaction between the catalytic groups, identifying their specific mechanistic roles. It transpired that in addition to functioning as the catalytic group, His57 could guide the attacking water toward the acyl group during deacylation. Other aspects of protein structure which are observable only by neutron diffraction analysis were also elucidated, including orientation of well-ordered amide side chains (made possible by the large scattering difference between nitrogen and oxygen atoms), the location and orientation of water molecules, and the hydrogen exchange properties of the protein. This example reveals the versatility of neutron diffraction when applied to the study of biomolecules, and neutron diffraction has been used to gain insight into many aspects of protein dynamics.

As we have seen, neutron diffraction clarified the serine protease mechanism, and may therefore be applied to similar mechanistic issues remaining to be resolved for the retroviral aspartic proteases. The difficulty in obtaining large enough crystals has until recently prevented successful neutron diffraction analysis of aspartic proteases. However, the development in recent years of the neutron Laue technique at the Intitut Laue-Langevin (ILL), and the European Molecular Biology Laboratory (EMBL) at Grenoble, has to some extent abrogated this problem. A neutron diffraction structure of the fungal aspartic protease endothiapepsin (Cooper and Myles 2000) is currently the largest protein solved by neutron diffraction methods, and has enabled the positions of the catalytic hydrogens to be established. It also represents the nearest analogy to the protonation states of the retroviral protease catalytic residues during hydrolysis. It is likely that similar experiments on retroviral protease crystals will eventually reveal the precise nature of their mechanism.

Coates *et al.* (Coates, Erskine, *et al.* 2001) collected high-resolution neutron diffraction data from crystals of the fungal aspartic proteinase endothiapepsin bound to a transition state analogue (H261). The functional groups at the catalytic centre had clearly undergone H-D exchange despite being buried by the inhibitor occupying the active site cleft. Crucially the data provided convincing evidence that Asp215 is protonated and that Asp32 is the negatively charged residue in the transition state complex. This analysis agreed with data used to propose an earlier mechanism for the enzyme (Veerapandian, Cooper, *et al.* 1992). However, the results also suggested that the hydrogen bonds between Asp32 and the hydroxyl of the transition state isostere are short - less than 2.6 Å – and thus possibly a particular type of hydrogen bond called a Low-Barrier Hydrogen Bond (LBHB). A brief summary of the LBHB concept can be found in the Appendix (Appendix: General Introduction - supplementary information; Low Barrier Hydrogen Bonds, page 339).

It has been proposed that LBHBs have an important effect in the catalytic mechanisms of a number of enzymes, including citrate synthase and serine proteinases. In serine proteases the formation of a LBHB follows steric compression resulting from the binding of the inhibitor. It is likely that a similar phenomenon occurs in aspartic proteases on inhibitor binding. The neutron diffraction data collected by Coates *et al.* strongly supports the position that Asp215 is protonated in the transition state, while their suggestion that a LBHB forms between Asp32 and the inhibitor hydroxyl would

account for delocalisation of the neutron density between these two groups. A subsequent 1.37 Å resolution crystal structure of endothiapepsin in complex with the gem-diol inhibitor PD-135,040 corroborated the previous endothiapepsin gem-diol inhibitor complex structures on which earlier catalytic mechanisms had been based. However, the increased resolution over previous structures also confirmed the presence of several short hydrogen bonds within the active site. Neutron diffraction studies of endothiapepsin are thus providing a valuable insight into the mechanism of aspartic proteases generally, and may finally answer the remaining questions in this field. Ultimately this will enable improvements in aspartic protease inhibitor design, and subsequent developments in HIV therapy.

HIV protease inhibitors.

Since the recognition of HIV PR as an essential component in the viral maturation process, and thus as a therapeutic target, a huge research effort has been applied to determine the enzyme's structural and mechanistic characteristics, and subsequently identify means by which the protease function may be blocked *in vivo*. Some aspects of protease inhibition have been briefly considered already, but it is now appropriate to consider the development of HIV protease inhibitors (and ultimately drugs) to date in more depth.

Prerequisites of drug design.

A candidate protease inhibitor molecule should posses certain characteristics. It must inhibit HIV protease and be resistant to processing itself, whilst ideally possessing a high binding affinity, high specificity, and a broad spectrum of activity against viral genotypes. It must not pose any risk to the health of a patient, and it must also demonstrate compatibility with other therapies, both primary anti-HIV therapies and secondary anti-opportunistic infection therapies. Compatibility with other anti-HIV therapies has become especially important following the development of combination therapies. Therapies to combat secondary infections may include the use of antibiotics, antivirals, or antifungals against opportunistic infections; vaccinations against infectious agents; and prophylaxis of Pneumocystis pneumonitis. New inhibitor compounds must have circulatory and cellular drug levels that are consistently greater than their inhibitory constant (K_i). They must exhibit good stability both *in vivo* (i.e. have long half-lives) and in terms of their formulation and storage. They must show good oral bioavailability, widespread tissue distribution, and be able to penetrate the blood brain barrier. Finally, where possible, they should be relatively cheap to manufacture and, in order to justify the high costs of drug development, it must be possible to protect the rights to any intellectual property involved - thus ensuring a return on investment for the developers. In reality a good inhibitor compound is likely to be a compromise solution that approaches each of these ideals most closely.

Testing activity against mutant strains.

HIV-1 exhibits a very high degree of genetic diversity (Domingo, Escarmis *et al.* 1996; Coffin 1992; Domingo and Holland 1997), which has been attributed to a high mutation rate (Mansky, Temin 1995), a rapid turnover (Perelson, Neumann *et al.* 1996; Wei, Gosh *et al.* 1995; Ho, Neumann *et al.* 1995), and a large viral population (Piatak, Saag *et al.* 1993; Embretson, Zupancic *et al.* 1993; Pantaleo, Graziosi *et al.* 1993; Chun, Carruth *et al.* 1997). The combined effect of these characteristics causes the relatively high frequency of mutations observed in HIV proteins. This in turn has resulted in the accelerated development of drug resistant HIV strains (Ho, Neumann *et al.* 1995; Wei, Ghosh *et al.* 1995; Condra, Schleif *et al.* 1995; Erickson and Burt 1996). Many mutations have been observed to occur in the HIV protease (Erickson, Gulnik *et al.* 1999), and in some cases these have conferred resistance to protease inhibitors. In addition, mutations may occur at the protease cleavage sites and in doing so compensate for any reduction in mutant protease efficiency (Cote, Brumme *et al.* 2001). Computer aided "co-evolution" studies have shown that compounds designed against the immutable properties (essential for the binding and cleavage of all native substrates) of the active site are better able to inhibit resistant strains, and also demonstrated that rational inhibitor design becomes more effective when designs are optimised against a large set of mutant enzymes (Rosin, Belew, *et al.* 1999). New compounds are now typically designed to evade mutants, and tested *in vitro* against a variety of the most commonly occurring protease mutants in order to maximise their efficacy *in vivo*.

Testing non-specific binding.

In vivo an inhibitor is exposed to a plethora of cellular and blood-borne proteins en route to its protease target, thus any tendency to bind these proteins non-specifically will reduce it efficacy. Candidate protease inhibitor compounds are therefore tested for non-specific binding using cell based assays in the presence of 50% human serum.

Testing bioavailability.

Bioavailability is a function of aqueous solubility and hepatic metabolism. Consequently it may be influenced by molecular size, and hydrogen bonding capability. In order to determine the degree of oral bioavailability a series of animal experiments are required. Candidate inhibitors are initially administered orally to rats, then to dogs and finally to monkeys. In each instance the compound is delivered at

pre-determined concentrations (mg / kg body weight) and subsequent levels of the compounds in the animal's blood are determined. The quantity detected in the blood may be influenced by the efficiency of absorption, or by the rate of drug metabolism.

Strategies for drug candidate isolation.

There are a number of possible strategies for identifying potential inhibitor molecules, such as random screening, rational screening, or rational design.

Random screening.

Random screening involves monitoring HIV protease activity, by means of a high throughput assay, in the presence of each molecule from a very large library of candidates. The candidates are not pre-selected in any way, and every available library of compounds is used in the screening process. When a candidate exhibits an inhibitory effect it is selected for further analysis. The screening process may then be biased towards candidates of a similar type in order to identify further leads, or the compound may form the basis for development of rationally designed molecules. Random screening requires a very large number of candidate molecules for screening; large quantities of active protease; and an easily monitored, rapid and reliable assay system. There is no guarantee that suitable lead molecules will be discovered, however, this is a widely used technique for drug discovery in general.

Rational screening.

Semi-random, or rational screening entails high throughput screens of candidates having structural similarities to known inhibitor molecules, or to leads derived from random screens.

Rational design.

Rational inhibitor design is an iterative process comprising *in silico* structure based design, combined with *in vitro* testing and structural solution. It relies heavily on prior structural data, and on the ability to obtain further structural data as designs progress. This imposes an additional requirement on potential compounds – their complexes with the target protease must crystallise to enable structural solutions to be obtained.

Each of these three general strategies has been employed in the search for HIV protease inhibitors, complementing and cross-fertilising each other in the process.

A review of HIV protease inhibitors and their development.

In considering inhibitor development it is useful to be familiar with the fundamentals of inhibitor-enzyme interaction and the basic strategies used to exploit these interactions. The enzyme's affinity for its substrate relies on a network of non-covalent bonds and interactions. The nature of these has been elucidated through the solution of early aspartic protease-inhibitor complex structures, and they have influenced the design of subsequent generations of inhibitor.

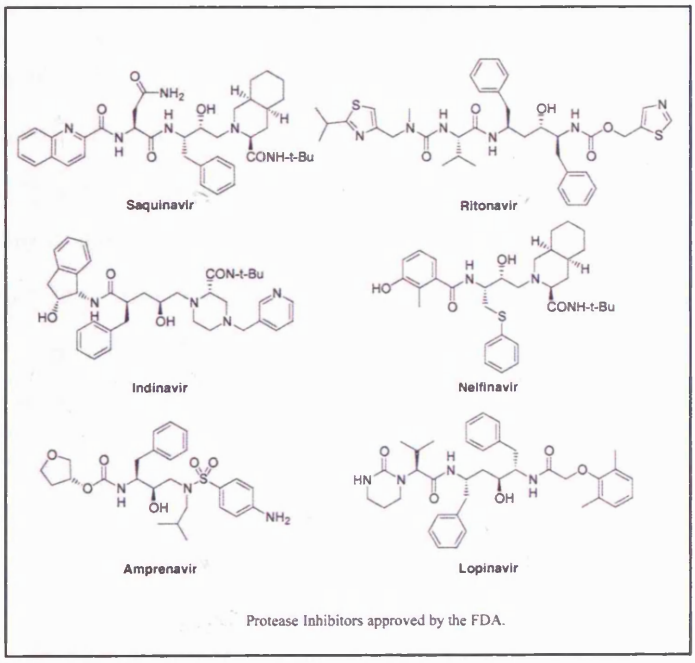
The most obviously influential interactions occur between the enzyme's substrate binding "pockets" and the inhibitor side chains. Inhibitor binding is mediated by numerous hydrogen bonds and Van der Waals contacts. In addition, inhibitors have frequently exploited hydrogen bonding between one or two inhibitor hydroxyl groups and the catalytic aspartates. These hydroxyls are intended to shift and replace the catalytic water molecule ordinarily coordinated by the catalytic aspartates, and responsible for nucleophilic attack on the substrate peptide bond. Hydrogen bonds also form between inhibitors and residues Gly27, Ala28, Ala29, Asp30, and Gly48, plus their alternate subunit counterparts (i.e. Gly27', Ala28', etc.) flanking the active site residues. A common feature revealed by structural studies of HIV PR-inhibitor complexes is a tetrahedrally coordinated water molecule linking the bound inhibitor molecule to the enzyme flaps (see fig. 25, page 97); noted in several structures it has been designated "water 301" (Swain, Miller, *et al.* 1990). This water accepts two hydrogen bonds from backbone amide hydrogen atoms of the of the flap residues Ile50 and Ile50' via its own oxygen atom, whilst donating two hydrogen bonds via it's hydrogen atoms to the carbonyl oxygen atoms of the inhibitor – thus inducing the fit of the flaps over the inhibitor. The distinction between the structural "water 301" - coordinated by the flaps and substrate - and the "nucleophilic" or "catalytic" water - coordinated by the active site aspartates (as proposed by the mechanism of Suguna, Padlan, *et al.* 1987; Pearl 1987) should be noted. Water 301 is located anterior to the scissile bond, while the nucleophilic water and the catalytic aspartyl side chain carboxyls are located on the posterior face of the scissile peptide bond (i.e. away from the flaps). In the following section the means by which these interactions have shaped inhibitor development will be considered, the problems associated with resistance will be considered and finally some alternative design strategies will also be explored.

Currently there are eight FDA approved HIV PR inhibitors compounds (table). The structures of six of these are shown in figure 19 below.

Table 1 FDA Approved HIV Protease Inhibitors.

FDA Approved Protease Inhibitors			
Brand Name	**Generic Name**	**Abbreviation**	**Manufacturer**
Agenerase	amprenavir	APV	GlaxoSmithKline
Crixivan	indinavir	IDV	Merck
Fortovase	saquinavir (soft gel cap)	SQV (SGC)	Roche
Invirase	saquinavir (hard gel cap)	SQV (HGC)	Roche
Kaletra	lopinavir/ritonavir	LPV/r	Abbott Laboratories
Lexiva	fosamprenavir	FPV	GlaxoSmithKline
Norvir	ritonavir	RTV	Abbott Laboratories
Reyataz	atazanavir	ATV	Bristol-Myers Squibb
Viracept	nelfinavir	NFV	Pfizer

Figure 19 Chemical structures of six FDA approved PR inhibitors.



Protease Inhibitors approved by the FDA.

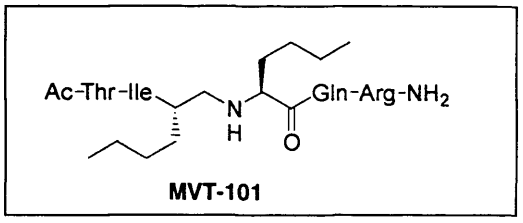
Non-hydrolysable peptide inhibitors.

Pepstatin A (or isovaleryl-Val-Val-Sta-Ala-Sta) with a K_i of 360 nM for HIV-1 PR, represented a starting point for inhibitor development, and Richards *et al.* were able to demonstrate that acetyl pepstatin had a K_i of 20 nM (Richards, Roberts *et al.* 1989). However, the pepstatins lacked specificity for the HIV PR, and were thus of little therapeutic use. Despite this, pepstatin, and other known inhibitors of aspartyl proteases, did provide an insight into the HIV protease mechanism - as discussed earlier in the chapter – and the statine moiety could also be incorporated into more specific inhibitor compounds, as will be seen.

The simplest strategy for the design of HIV protease specific inhibitors was to base them on a peptide substrate. This was facilitated by the characterisation of substrate-protease interactions that resulted in cleavage (considered on pages 59-64 this chapter). Studies by Griffiths *et al.* (Griffiths, Phylip *et al.* 1992) found a preference for proteolysis of substrates with scissile hydrophobic-hydrophobic (usually involving leucine, alanine, or valine) or aromatic-proline (i.e. phenylalanine-proline) peptide bonds. The substitution of a non-cleavable bond for the scissile site generates an inhibitor peptide.

Many of the earliest non-scissile substrate based inhibitors incorporated a reduced amide bond (Moore, Bryan *et al.* 1989), and the first protease-inhibitor complex structure solved by crystallography was that of HIV-1 protease in complex with the reduced amide inhibitor MVT-101 (fig. 20) (Miller, Schneider, *et al.* 1989). This compound did not have the central hydroxyl group that would become a feature of many subsequent inhibitors, consequently it did not interact with the catalytic aspartates and suffered low affinity.

Figure 20 The reduced amide inhibitor MVT-101.



Structural analyses of peptide based compounds revealed a large degree of hydrogen bonding between the enzyme and peptide backbone suggesting relatively non-specific, promiscuous, peptide binding characteristics. Consequently the variation of amino acid residues proved an unsuccessful strategy for the design of peptide based HIV PR specific inhibitors. In any case reduced amide compounds exhibited relatively low potency, and peptide based compounds displayed very poor pharmacokinetic properties (Plattner and Norbeck 1989).

Transition state analogue inhibitors.

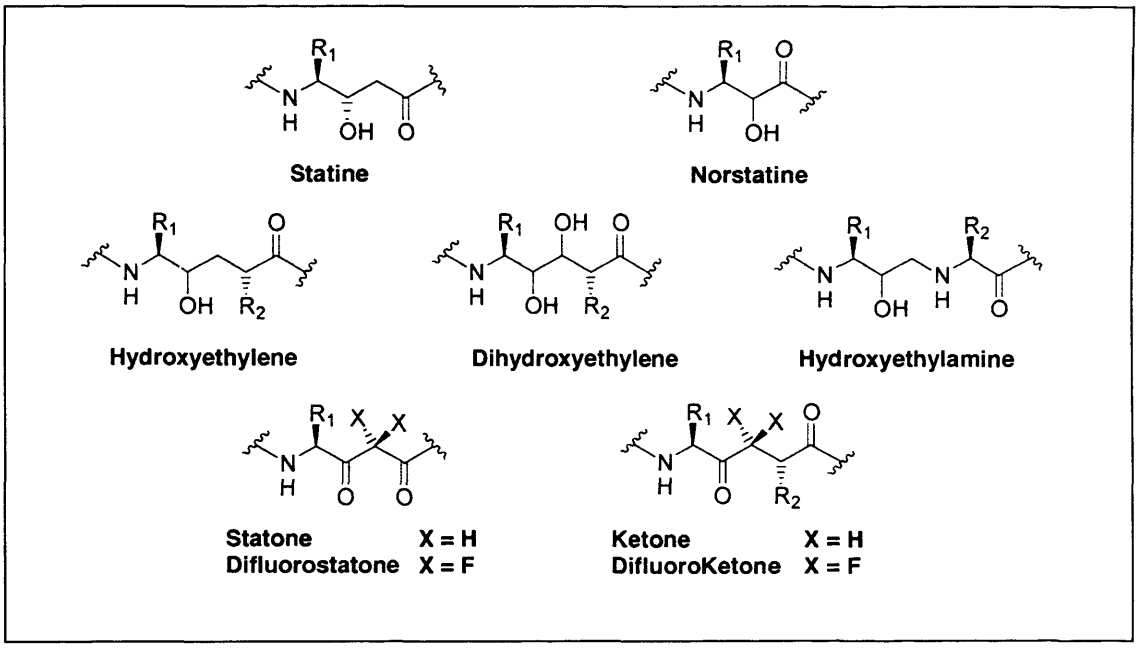
Since the non-hydrolysable peptide isosteres - such as the reduced amides - had generally failed to provide effective therapeutic compounds, a mechanism based inhibitor design strategy was now pursued. Research into renin inhibitors (see Appendix, pages 332-333) had shown that transition state analogues were bound very tightly. The affinity of a substrate to an enzyme is generally very much lower than that of its transition state (Wolfenden 1999), thus the potency of protease inhibitors could be increased if they were based instead on the substrate transition state. Studies of renin had revealed the tetrahedral intermediate as approximating to the aspartic proteinase transition state, and identified statine, norstatine, hydroxyethylene and hydroxyethylamine based compounds as analogues of this state (fig. 21).

In 1989 Richards reported that compound H-261 (tBoc-His-Pro-Phe-His-Leu-ψ[CH(OH)]CH2]-Val-Ile-His[5]), known to be a potent inhibitor of human renin (Blundell, Cooper *et al.* 1987; Leckie, Szelke *et al.* 1985) and most typical aspartic proteinases (Valler, Kay *et al.* 1985), had a K_i of 5 nM with respect to HIV-1 PR at pH 4.7 (Richards, Roberts *et al.* 1989). This hydroxyethylene compound was non-scissile and mimicked the sp3 carbon hybridisation of the supposed transition state geometry. However, it showed no specificity for HIV PR over other aspartic

[5] (ψ[CH(OH)CH2] indicates the hydroxy-ethyl peptide bond replacement)

proteases, such as human renin. Observations such as those by Richards *et al.* ensured that advances made during the development of transition state based renin inhibitors (Szelke, Leckie *et al.* 1982; Rich 1990; Wiley and Rich 1993) would subsequently be applied to the design of HIV protease inhibitors. The development of renin inhibitors is considered further in the appendix (page 331-332).

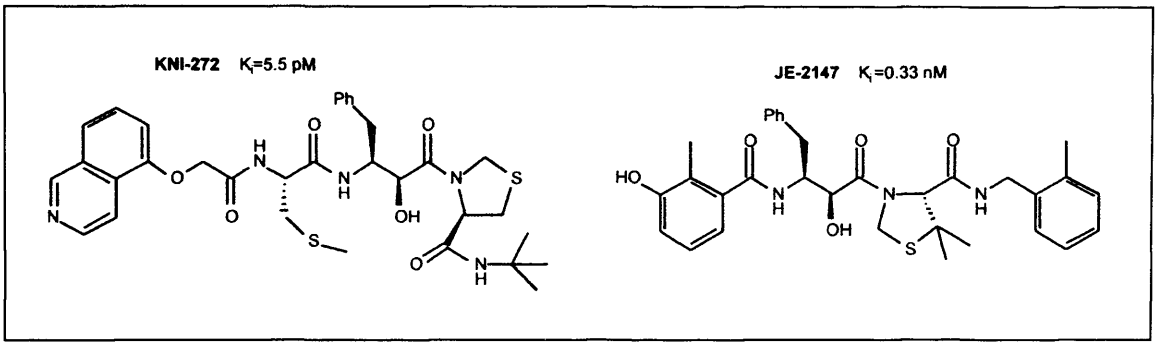
Figure 21 The functional moieties of transition state analogue inhibitors.



In pursuit of transition state analogue inhibitors of HIV protease, the statine and norstatine moieties were incorporated into substrate-analogue peptides (Raju and Deshpande 1991; Fehrentz, Chomier *et al.* 1992; Sakurai, Sugano *et al.* 1993; Mimoto, Imai, *et al.* 1991; Tam, Carriere *et al.* 1992; Slee, Laslo *et al.* 1995). Statine based compounds displayed low potency, probably because in lacking a P1′ side-chain the number of hydrophobic contacts between compound and PR were reduced, lowering affinity for the enzyme (Sakurai, Sugano *et al.* 1993; Kempf 1994). In contrast the norstatine based compounds showed greater activity since they more closely resembled the native substrate. Norstatine has one (-CH_2-) group less than statine and allows the reintroduction of a normal P1' residue. Compound KNI-272 (fig. 22 left) was based on norstatine and proved to be a potent HIV-1 PR inhibitor (Baldwin, Bhat, *et al.* 1995 (1); Baldwin, Bhat, *et al.* 1995 (2); Kiso 1996). However, despite low toxicity, it displayed poor bioavailability and was rapidly metabolised (Chokekijchai, Shirasaka, *et al.* 1995; Kiriyama, Nishiura *et al.* 1996; Mueller, Anderson, *et al.* 1998), whilst HIV-1 quickly developed resistant mutants when

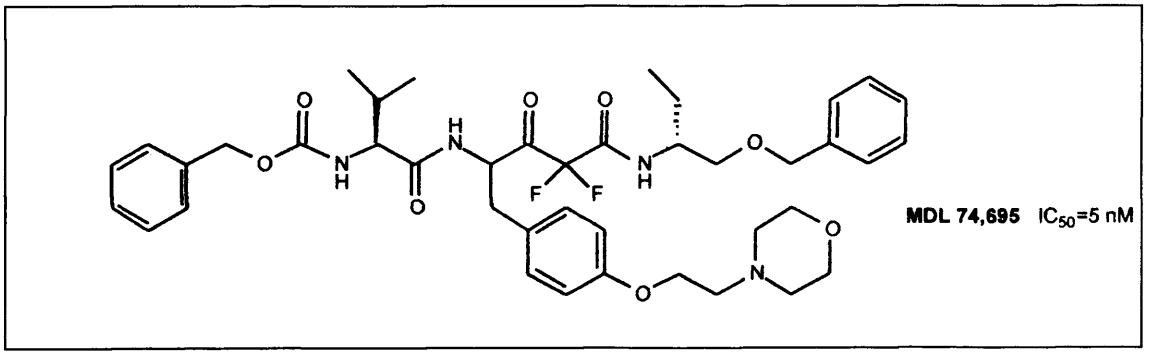
exposed to the compound *in vivo*. KNI-272 was subsequently used as the basis for JE-2147 (fig. 22 right), a potent norstatine based inhibitor incorporating a cyclic P2 moiety that appeared to improve pharmacokinetic properties (Mimoto, Kato, *et al.* 1999; Kiso, Yamaguchi, *et al.* 1998).

Figure 22 The Norstatine based inhibitors KN-272 and JE-2147.



Difluoroketones (Sham, Betebenner *et al.* 1991; Sham, Betebenner *et al.* 1993) and difluorostatones have also been investigated as HIV PR inhibitors, employing a central fluorinated ketone moiety (fig. 21) to mimic the tetrahedral intermediate. Among examples of such inhibitors was compound MDL 74,695 (fig. 23) which had an IC50 of 5 nM, but unfortunately exhibited poor bioavailability, and rapidly selected for resistant mutants *in vivo*.

Figure 23 The Fluorinated ketone inhibitor MDL 74,695.



The crystal structure of acetyl-pepstatin complexed with HIV protease was solved at 2.0-2.5 Å resolution (Fitzgerald, McKeever *et al.* 1990) and demonstrated that the pattern of hydrogen bonding and subsite interactions for acetyl-pepstatin was similar to those observed for the hydroxyethylene inhibitor-HIV PR complexes. The main difference was the lack of a P1' residue occupying the S1' subsite. The similarity between statine and hydroxyethylene based inhibitors is unsurprising when it is

considered that statines are effectively truncated hydroxyethylene transition isosteres of Leu-Gly where the P1 amide nitrogen is absent. However, their symmetrically hydrogen-bonded P1 and P2' amide nitrogens are separated by only five bonds as opposed to the six present in peptide substrates, or hydroxyethylene and reduced amide isosteres.

Hydroxyethylene based compounds proved more potent than their statine or norstatine alternatives, and were the first compounds to exhibit activity *in vivo*, using cell-based assays (Dreyer, Metcalf *et al.* 1989; Meek, Lambert *et al.* 1990; Dreyer, Lambert *et al.* 1992; Roberts, Martin *et al.* 1990). Structures of HIV-1 protease complexed with hydroxyethylene-containing peptide mimics (Jaskolski, Tomasselli *et al.* 1991; Dreyer, Lambert *et al.* 1992; Graves, Hatada *et al.* 1992; Thompson, Fitzgerald *et al.* 1992) have shown that the hydroxyl group on the inhibitor adopts, at least approximately, the position usually occupied by the active site water molecule. Hydrogen bonding analysis of these hydroxyethylene inhibitor complexes lead to the conclusion that one of the two aspartates must be protonated in order to bind to the inhibitor hydroxyl group.

The majority of published hydroxyethylene based compounds have exhibited an *S* conformation at their hydroxyl group. The explanation for this becomes clear when it is considered that it is the stereochemical configuration of the hydroxyl group that determines the degree of interaction with the active site carboxylate side chains. In the *S* conformation (where the *S* hydroxyl group is in anti configuration with respect to the preceding *R* group of the P1 substituent) the hydroxyl group is positioned such that it interacts with both carboxylates, whereas in the *R* conformation similar interactions would require major conformational rearrangements in the inhibitor backbone. Consequently a hydroxyethylene inhibitor compound having an *R* hydroxyl would adopt a higher energy inhibitor conformation, or exhibit weaker non-bonded interactions with the enzyme, or both it would therefore be less effective as an inhibitor compound.

Dihydroxyethylene groups have also been used to generate HIV PR inhibitors, such as U-75875 (fig. 24), which have a high affinity for the protease (Ashorn, McQuade *et al.* 1990). Structural studies of the protease complexed with U-75875 indicated that

the N-terminal hydroxyl group interacts equally well with either active site aspartate and in a manner similar to the isosteric *S* hydroxyl group of the monohydroxy inhibitors (Thanki, Rao, *et al.* 1992). Meanwhile the second hydroxyl group was only able to hydrogen bond with one of the aspartates (Asp 25), so that its contribution to inhibitor potency remains unclear, and the benefits derived from inclusion of an additional hydroxyl group appear negligible (Kempf and Sham 1996).

Figure 24 Structure of the dihydroxyethylene based inhibitor U-75875.

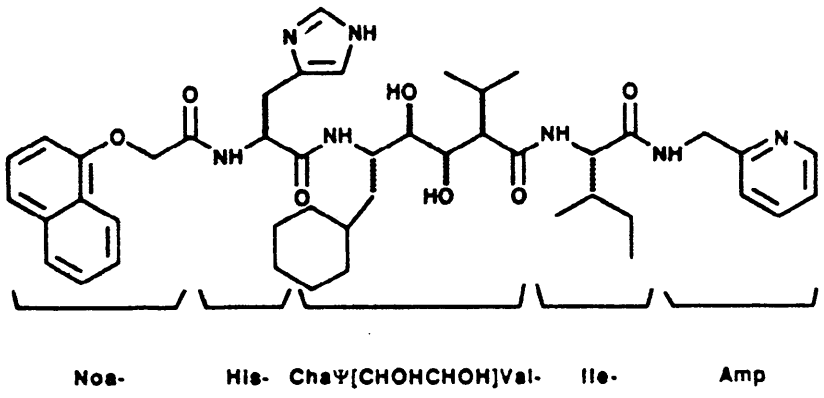
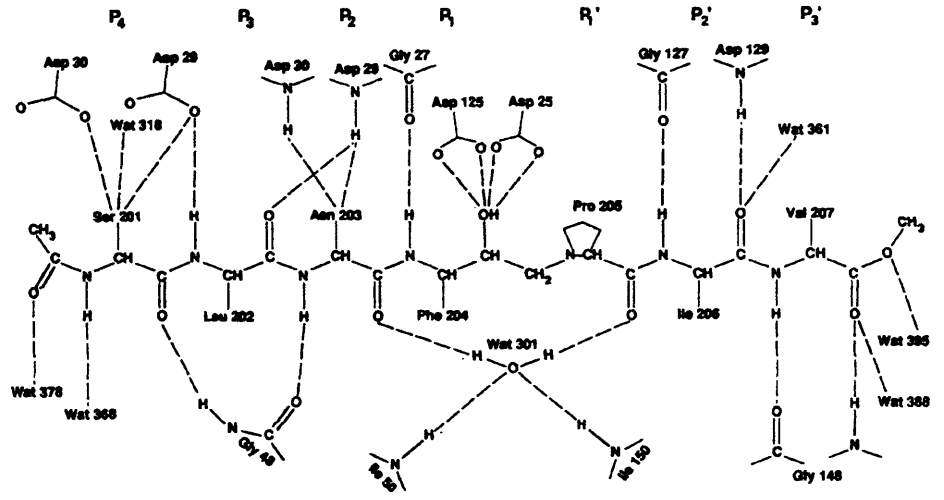


Figure taken from Ashorn, McQuade, *et al.* 1990

The hydroxyethylene compounds had tended to suffer from poor aqueous solubility and oral bioavailability, consequently the incorporation of a basic amine into their backbone was considered as a means to improve their characteristics. This resulted in development of the hydroxyethylamines.

Hydroxyethylamine based compounds introduce an additional bond between the P1 and P1' residues in comparison to the reduced amide and hydroxyethylene based compounds that had preceded them. Nevertheless, structural studies of HIV PR complexed with JG-365, a heptapeptide [CH(OH)-CH2-N] analogue (fig. 25), demonstrated that the inhibitor's backbone adopted a similar conformation to that observed for both the reduced amide and hydroxyethylene based inhibitor structures (Swain, Miller *et al.* 1990), despite the additional bond and the constrained prolyl group. The JG-365 preparation was a racemic mixture of the *R* and *S* configurations at the hydroxyl carbon; however, the *S* isomer was exclusively present within the crystal structure. Furthermore the *S* hydroxyl group could be superimposed onto the *S* hydroxyl groups of hydroxyethylene based inhibitors, demonstrating that they adopted the same conformation.

Figure 25 The dihydroxyethylamine inhibitor JG-365 binding HIV-1 PR.



Hydrogen bonding interactions between HIV-1 PR and JG-365 are represented by dashed lines.

Figure taken from Swain, Miller, *et al.* 1990.

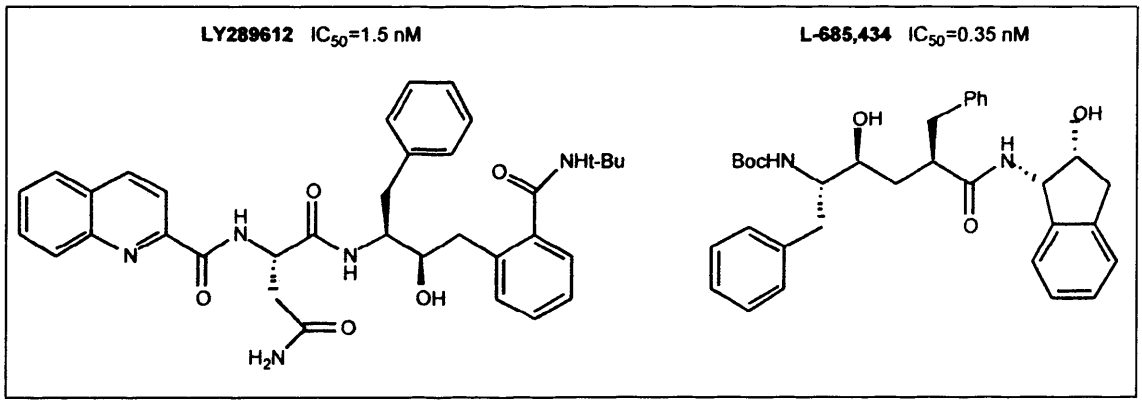
In 1990 Roberts *et al.* from Roche chose the Phe-Pro cleavage site sequence, located between the RT and PR of Pol, as their target (Roberts, Martin *et al.* 1990). By shortening the inhibitory peptide from six residues - initially believed to be the requisite minimum - to three, and by incorporating a hydroxyethylamino isosteric replacement, they generated a series of test compounds. Of these the compound Ro-31-8959 (Roberts, Martin *et al.* 1990), later known as Invirase or Saquinavir (see fig. 19), was shown to be an effective inhibitor and became the first HIV protease inhibitor to be administered to humans. The compound consisted of a quinoline, asparagine, phenylalanine, decahydro-isoquinoline (DIQ) and NH-tertiary butyl. The K_i of this hydroxyethylamino isostere against HIV-1 PR was 0.4 nM, with antiviral activity in the nanomolar range. In comparison, the RT inhibitor (and first approved anti-HIV therapy) AZT was not active at 100 μM. Consequently this new anti-PR compound represented a breakthrough in AIDS therapy, and Saquinavir became the first clinically approved HIV protease inhibitor.

The shortened inhibitor peptides employed by Roberts *et al.* lacked a P3' residue, this revealed that the R, or *syn*, hydroxyl group configuration is preferred by hydroxyethylamine based inhibitors without a P3' residue. This contrasted with the preference for the *anti* hydroxyl group configuration observed in all previously

described hydroxyethylene based protease inhibitors (both HIV PR and other aspartic protease peptidomimetic inhibitors). The *R* hydroxyl group compounds therefore differed from the *S* hydroxyl group compounds in their mode of enzyme binding. The binding mode for Ro-31-8959 with HIV protease was determined crystallographically (Krohn, Redshaw *et al.* 1991) and, as had been predicted, the DIQ and *t*-butyl groups bound the S1' and S2' sub-sites, respectively. The DIQ carbonyl group was hydrogen bonded via the buried water to the flaps. However, the Nitrogen atom of the DIQ moiety had the opposite configuration to that of the corresponding proline ring in the predecessor hydroxyethylamine compound JG-365, whilst the nitrogen of the t-butylamide was displaced by approximately 1.8 Å preventing it from hydrogen bonding with the Gly27 carbonyl oxygen, as it might otherwise have done. The inhibitor backbone position prohibited any future chain extension into the S3' subsite, and steric hindrance by the large DIQ group in the P1' position virtually precluded a hydroxyl group in the *S* configuration - unlike compounds having a proline at the P1' site - and explaining the preference for an *R* configuration.

The development of hydroxyethylamines lead to reappraisal of some preceding hydroxyethylene compounds – in particular LY289612 and L-685, 434 (fig. 26), both designed to imitate a Phe-Pro cleavage site.

Figure 26 The hydroxyethylamine inhibitors LY289612 and L-685,434.



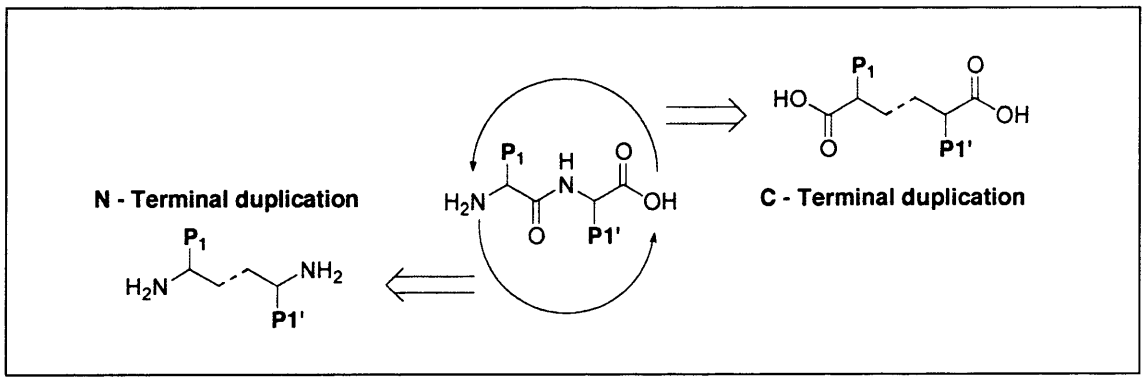
The incorporation of a basic amine into their backbones resulted in significant improvements and ultimately lead to their evolution into the FDA approved drugs Nelfinavir and Indinavir shown in figure 19 (Kaldor, Kalish *et al.* 1997; Reich, Melnick, *et al.* 1995; Jungheim, Shepherd, *et al.* 1996). Nelfinavir evolved from designs based on both Saquinavir and LY289612, each of these predecessors having exhibited poor oral bioavailability (Kaldor, Kalish *et al.* 1997; Reich, Melnick, *et al.*

1996; Melnick, Reich, *et al.* 1996). Improvements were achieved by incorporation of an S-phenyl cysteine spanning P1-P3, and thereby improving hydrophobic interaction with the enzyme at these positions. The intervening P2 group was then optimised accordingly, and resulted in AG1343, also known as Nelfinavir. Similarly the introduction of an amine into L-685, 434, was intended to improve aqueous solubility and oral bioavailability (Thompson, Murthy, *et al.* 1994) and lead to the optimisation of several related hydroxyaminopentane amides (Vacca, Dorsey *et al.* 1994; Chen, Li *et al.* 1994; Dorsey, Levin, *et al.* 1994), culminating in L735,524 - a compound displaying good pharmacokinetic properties and oral bioavailability, later known as Indinavir.

Symmetrical and pseudo-symmetrical peptidomimetic inhibitors.

Structural solutions of HIV protease in complex with asymmetric (peptide based) inhibitors revealed that both their N and C termini bound identical subsites as a result of the enzyme's C_2 symmetry. The design strategy of N or C terminal duplication (fig. 27) resulted from this observation, and the development of symmetrical and pseudo-symmetrical inhibitor compounds followed. One of the advantages anticipated of this approach was high selectivity for the HIV PR; since the viral enzyme's symmetrical nature was unique. This approach was initially adopted by Erickson *et al.* (Erickson, Neidhart, *et al.* 1990) who started with known inhibitors such as D-His-Pro-Phe-His-Phe-ψ[CH2NH]-Phe-Val-Tyr and made use of the complex structure of the Rous Sarcoma Virus protease. They exploited the C2 symmetry of this protease by replicating the left hand side 'N-terminal' of their peptide mimic on the right hand side, and incorporating an almost symmetrical tertiary hydroxyl moiety between them.

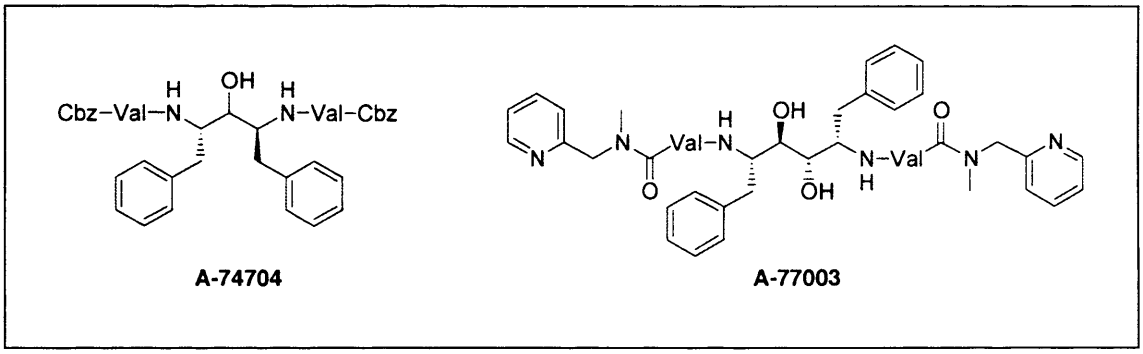
Figure 27 The N and C terminal duplication inhibitor design strategies.



Effectively they produced a peptidomimetic with twin N-termini. The compound had a K_i of 4 nM, and was 10000 times more active against HIV PR than against other

proteases. An equally feasible alternative involved replication of the C-terminus at each end. The N or C terminal duplication of inhibitors was intended to confer them with greater potency and specificity. The N-terminal duplication strategy was used by Abbot to generate the pseudo-symmetric di-amino-alcohol core of inhibitor A-74704 (fig. 28, left), and X-ray crystallography confirmed highly symmetric binding of this inhibitor to the HIV PR active site. Abbott used the same strategy to develop the di-amino-diol core, as in compound A-77003 (fig. 28, right) – which was clinically tested as an intravenous drug (Reedijk, Boucher *et al.* 1995).

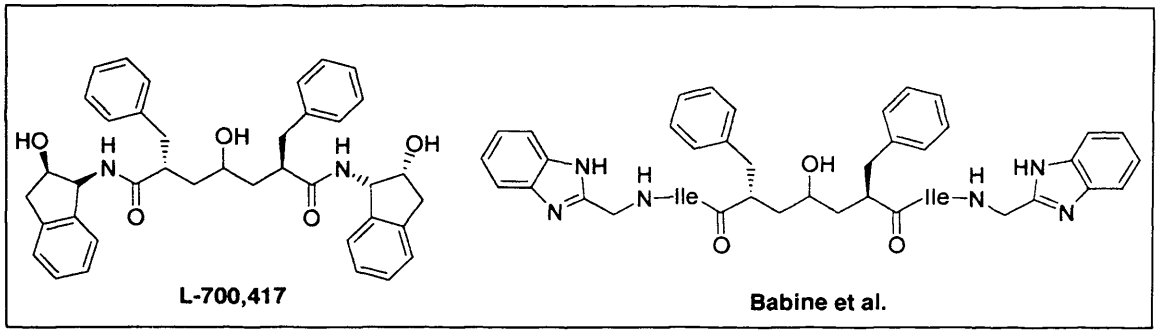
Figure 28 The N terminally duplicated inhibitors A-74704 and A-77003.



Rationally designed improvements were made to this class of compounds, including a systematic investigation of P3 and P2' heterocyclic groups intended to decrease the rate of hepatic metabolism. The eventual replacement of the pyridyl groups with thiazoles provided increased chemical stability towards oxidation whilst maintaining sufficient aqueous solubility for oral absorption. These modifications generated the non-symmetric clinically approved compound Ritonavir (Erickson and Kempf 1994; Kempf, Marsh *et al.* 1995; Kempf, Sham *et al.* 1998).

The N-terminal duplication strategy has also been used by several other groups (Budt, Peyman *et al.* 1995; Mo, Markowitz *et al.* 1996, Ettmayer, Hubner *et al.* 1994). The C-terminal duplication strategy has also been employed to devise inhibitors (Kempf 1994), but to a lesser extent. Bone *et al.* were the first to report a high affinity inhibitor, L-700,417 (fig. 29, left), exploiting the C-terminal duplication strategy, and crystallographic analysis revealed highly symmetric binding (Bone, Vacca *et al.* 1991).

Figure 29 Other terminally duplicated peptidomimetic inhibitors.



Some incorporate one or two hydroxyl groups, maintaining a fully symmetric or quasi-symmetric character, while others are phosphinate based (Abdel-Meguid *et al.* 1993). Their mode of binding does not necessarily reflect their structural symmetry, and some have been shown to bind asymmetrically (Dreyer, Boehm *et al.* 1993).

A more peptide-like inhibitor (fig. 29, right), which also exhibited a relatively high potency, was designed and synthesised independently by Babine *et al.* (Babine, Zhang *et al.* 1993). In designing these peptidomimetic compounds the hydrogen bonding and electrostatic interactions along the cylindrical groove, and optimal fitting of amino acid side chains into the binding groove were taken into account. Also, their central carbonyl groups were designed to interact, via hydrogen bonding, with water molecule 301 in order to stabilize the close formation of the flaps on the enzyme (West and Fairlie 1995). While inhibitors such as Ro31-8959, LY289612, and other similar peptidomimetic protease inhibitors proved to be potent antivirals *in vitro,* they suffered from poor absorption, poor oral bioavailability, short serum half-life values, high susceptibility to hydrolysis by degradative enzymes, and other pharmacological problems (Reich, Melnick *et al.* 1995). The limitations on peptidomimetic inhibitors have resulted in efforts to minimise the pharmacological problems by developing inhibitors that are structural mimics of peptides but with little or no peptidic character.

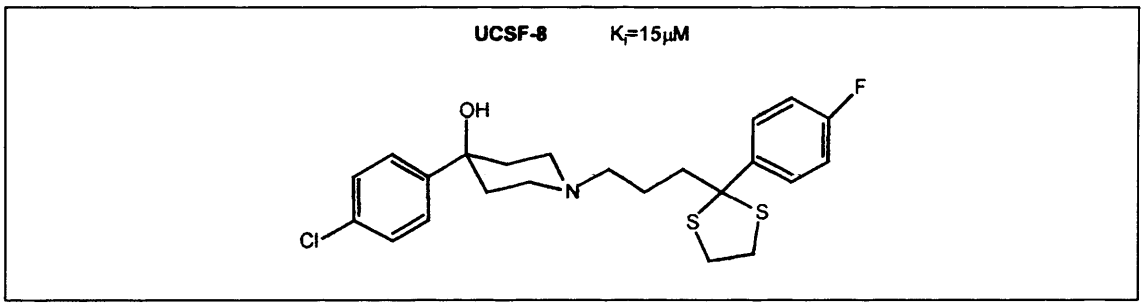
Reich, Melnick *et al.* (1995) approached this problem by beginning with the co-crystal structure of HIV-1 protease with LY289612 (a peptidomimetic inhibitor). As their base design they retained only the non-peptidic hydroxyethyl-t-butylbenzamide portion of the original inhibitor. They then synthesized and attached a variety of substituents to fit the various active site pockets. An iterative, seven-cycle, process

involving solving the co-crystal structures of the protease with each new non-peptide inhibitor, and using the data to design the next, ultimately resulted in improved design and maximised binding affinities. Further characterization *in vivo* showed improved bioavailability of up to 30% and maintenance in plasma levels, at or above, the antiviral IC90 for five or more hours in rats, dogs, and monkeys. However, the majority of peptidomimetics exhibited poor bioavailability due to their inherent low solubility, and high clearance. Consequently the development of non-peptide inhibitor designs was preferred.

Non-peptide inhibitors.

Non-peptidic inhibitors included the compound UCSF-8 (Rutenber, Fauman *et al.* 1993), a modification of the drug haloperidol. Haloperidol was discovered to be an inhibitor of HIV-1 PR following a computational screen of the 10000 molecules comprising the Cambridge Structural Database. The screen sought those molecules that were complimentary to the HIV-1 PR active site in terms of their steric properties. However, haloperidol is toxic at high concentrations and thus unsuitable as a drug. Replacement of haloperidol's functional ketone with a thioketal ring resulted in compound USCF8 shown in figure 30 (Rutenberg, Fauman *et al.* 1993).

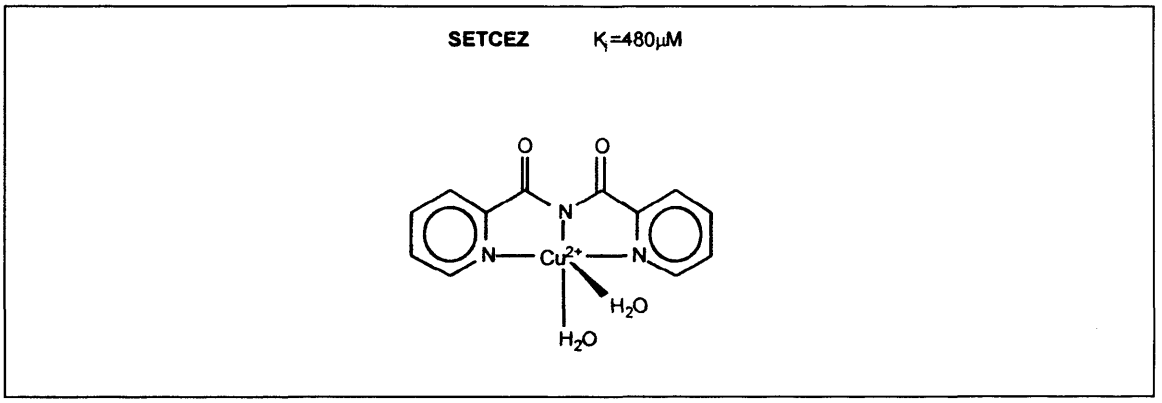
Figure 30 The haloperidol based inhibitor UCSF-8.



In contrast to peptidomimetic inhibitors, USCF8 interacts mainly with the flaps of the enzyme in an open conformation. The interaction of USCF8 with the catalytic site of the protease is mediated by a water molecule located at equal distance (3 Å) from two catalytic aspartates. This water molecule probably occupies a similar position to the catalytic water molecule responsible for substrate hydrolysis, but since the peptidomimetic inhibitors replace this water molecule with their hydroxyl groups, it had not previously been observed in their crystallographic structures. A

pharmacophore[6] was subsequently designed taking this water into account, and the resultant model - which addresses metal-organic compounds particularly - was used for screening new HIV PR inhibitors through the Cambridge Structural Database. Among the structures fitting into the active site of the enzyme, the compound SETCEZ, or Diaqua [bis (2-pyridylcarbonyl) amido] copper (II) Nitrate (Castro, Faus, *et al.* 1990), shown in figure 31, was found to behave as a competitive inhibitor of HIV-1 protease with a K_i value of 480 uM (Lebon and Ledecq 2000). New copper coordination compounds, having an octahedral geometry favorable for the orientation of their interacting substituents within the protease subsites were designed and inhibited HIV-1 protease in the micromolar range (Lebon and Ledecq 2000).

Figure 31 The organo-metallic inhibitor SETCEZ.



[6] The traditional medicinal chemistry definition of a pharmacophore is the minimum functionality a molecule has to contain in order to exhibit activity. Only molecules which interact at the same receptor site in the same way will share a pharmacophore.

The significance of "water 301".

As structural data became more readily available, and the use of modelling and simulations became more common, structure based design prevailed. Molecular modelling eventually became an essential aid to crystallographers and medicinal chemists in the design of HIV protease inhibitors. Computer based designs enabled the development of non-peptide inhibitors able to displace both the catalytic water, and the structural water molecule 301, whilst exhibiting much improved pharmacokinetic properties.

Water 301 acts as a bridge between substrate and the enzyme flaps. It has been suggested that the hydrogen bonds contribute to distortion of the scissile bond away from planarity, increasing the electrophilic character of the bond carbonyl and facilitating catalysis of the bonds cleavage by the nucleophilic water. Baca and Kent showed that specific deletion of just these two hydrogen bonds leads to a drastic decrease in the catalytic efficiency of the enzyme (Baca and Kent 1993). This coordinated structural water (water 301) is found exclusively in the retroviral aspartic proteinases, and is not observed in the eukaryotic aspartic proteinases. Consequently the contrasting structural observations, and anticipated mechanistic differences between retroviral and eukaryotic aspartic proteinases relating to water 301, have provided a focus for structure-based drug design targeted at retroviral proteinases. The eukaryotic, pepsin-like, aspartic proteinases are two-domain, single polypeptide chain (i.e. monomeric) molecules having only one flap. Their single flap interacts with the substrate P2 and P10 carbonyls by direct, non-water-mediated, hydrogen bonding. Inhibitor designs have sought to exploit this difference in enzyme-inhibitor H-bonding between the aspartyl proteinases of human host and parasitic retrovirus respectively, based on the presumption that it is mechanistically relevant. The intention has been to generate inhibitors specific for HIV PR that have no deleterious effect on the function of host cell-encoded aspartyl proteinases.

The cyclic urea inhibitors.

Lam *et al.* from DuPont Merck were among the first to pursue this approach, publishing the rational design of their non-peptidic cyclic urea inhibitors (XM323 and XM412) in 1994 (Lam, Jadhav *et al.* 1994). Subsequent crystallographic analyses confirmed the computer generated predictions of successful water displacement by these compounds.

Figure 32 The design evolution of cyclic urea inhibitors.

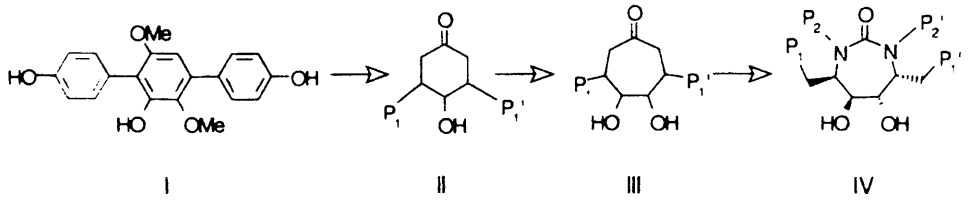
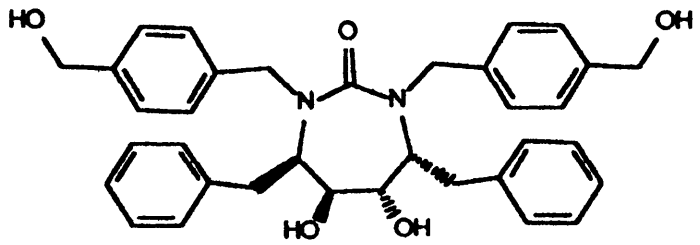


Figure taken from Hodge, Straatsma *et al.* 1998, and based on work by Lam, Jadhav *et al.* 1994.

Lam *et al.* began by designing a C2-symmetric dihydroxy-ethylene mimic, and then using the Docking molecular modelling process to locate it within a protease structure. They then proposed a simple pharmacophore based on a hydrogen bonding group near the active site aspartates and two symmetric hydrophobic groups for P1 and P1'. Three dimensional database searching presented the 1,6-para-diphenol-2,5-dimethoxy-6-hydroxybenzene molecule (fig. 32 I), where the phenol groups occupied the P1 and P1' positions. The methoxy and hydroxyl substituents were designed to form hydrogen bonds within the active site, with the 2-methoxy group specifically intended to displace water molecule 301. For better positioning of these groups they selected a cyclohexanone ring (fig. 32 II), then expanded this to a seven-membered ring to accommodate a diol (fig. 32 III). Finally they introduced a cylic urea (fig. 32 IV) to improve hydrogen bonding with the flap isoleucines. Additional modelling predicted the stereochemistry and conformation which was later confirmed by an X-ray structure. The resultant compound was a cyclic urea peptidomimetic of D-phenylalanines. Two further small hydrophobic groups for the P2 and P2' site (represented by P2 and P2' in fig. 32 IV) were produced by N-substitution, reducing the K_i from 4 μM to 4 nM. The hydrophobic groups, the hydrogen bonding pattern and the rigidity combined to make these compounds good inhibitors. After modelling in the S2 and S2' site, the hydrophobic side chains were enlarged to improve the

inhibition to 0.3 nM. A hydrogen bonding group - OH, in XM323 (fig. 33), and later NH2 in XM412 - on the phenyl rings, improved the antiviral activity, and the bioavailability. Subsequent crystal structures have seemed to confirm the proposed binding mode (see fig. 34). While XM323 has 50% bioavailability and drug levels in plasma exceeds the IC90 for 4 to 16 hours, Phase I clinical trials were abandoned in 1993 in favour of the derivative XM412 which was found to be more water soluble and bio-available (West and Fairlie 1995).

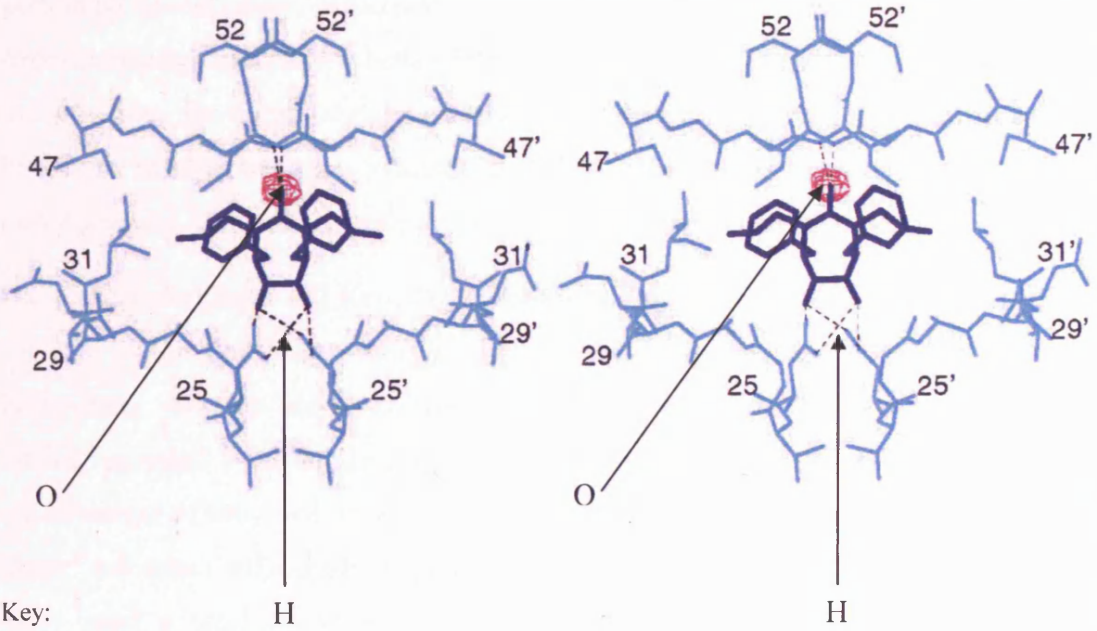
Figure 33 The cyclic urea inhibitor candidate XM323



XM323

Figure taken from Otto, Reid *et al.* 1993, and based on work by Lam, Jadhav *et al.* 1994

Figure 34 Stereo image of a cyclic urea (XK216) displacing water 301.



Key:

--- = hydrogen bonds

O = position of inhibitor oxygen (original position of water 301) and hydrogen bonds to flap

H = Hydrogen bonds between inhibitor and catalytic Aspartates.

Figure adapted from Ala, DeLoskey *et al.* 1998, showing HIV-1 PR (light blue) complexed with cyclic urea XK216 (dark blue). Water 301 has been displaced by the inhibitor oxygen, which also interacts with the flaps. The original position of water 301 prior to displacement is marked by superimposed electron density (pink).

Given the differences between retroviral and host aspartyl proteases despite their presumed evolutionary, structural, and mechanistic similarities it was important to experimentally determine whether HIV PR required functional H-bonding from one or both flaps for normal catalytic activity. As already described the water mediated hydrogen bonding between substrate and flaps appeared to have important catalytic consequences; it was therefore a significant consideration in drug design.

Baca and Kent (Baca and Kent 1993, Baca and Kent 2000) determined to investigate the role of the flaps and water 301 in catalysis. As earlier described, HIV PR is a homodimer of two identical, non-covalently associated, 99-residue polypeptide chains, encoded by a single gene. Consequently, in order to carry out asymmetric modification of only one of the two monomers, Baca and Kent produced a "tethered dimer" construct, effectively making a single polypeptide of the two monomer chains. They used a total chemical synthesis technique, involving convergent chemical ligation, to generate a 202-residue synthetic HIV-1 protease whereby the two subunits were covalently joined by means of a short linker structure between the C-terminus of one and the N-terminus of the other.

Previously, covalent dimer analogues of HIV-1 PR had similarly been produced by recombinant DNA expression in bacterial systems and were shown to be identical to the native enzyme in terms of their kinetic properties (DiIanni, Davis *et al.* 1990; Cheng, Yin, *et al.* 1990; Tozser, Yin, *et al.* 1997). The engineered, synthetic, HIV-1 PR analogue was generated such the Gly49-Ile50 amide bond N(H) was replaced in one flap only by an oxygen atom not capable of acting as a hydrogen bond donor. A control enzyme was also synthesised in which no such modification was introduced.

If, as predicted based on crystallographic observations, each flap contributed a catalytically relevant hydrogen bond then a PR analogue engineered to eliminate this H-bonding in one flap only would be expected to exhibit significantly reduced catalytic efficiency. Previously Baca and Kent had performed experiments in which they had used a similar technique to delete the relevant H-bonding in both flaps simultaneously by incorporation of sulphur to replace each of the Gly49-Ile50 backbone amides (Baca and Kent 1993). This earlier experiment had resulted in considerable loss of activity, and based on their observations it was predicted that an asymmetrically engineered construct would display a rate (kcat) reduction of 100-fold

compared with native HIV-1 PR. It was anticipated that the kcat reduction would result from a combination of two effects: an increase in the activation energy of enzyme-catalyzed substrate hydrolysis, and a two-fold rate reduction due to the asymmetry of the covalent dimer form of the enzyme - which would permit only one of the usual two orientations of substrate binding. However, if both the enzyme-substrate P2-P1' carbonyl H-bond(s) can be donated from a single flap (as in the pepsin-like proteinases), no change in intrinsic catalytic activity would result from deletion of the essential H-bonding element in only one of the two flaps. In this event, only a two-fold rate (kcat) reduction would be expected (because only one productive substrate binding mode would be possible in the now-asymmetric enzyme).

The catalytic activity and substrate specificity of both the Control and Modified HIV-1 PR constructs were evaluated by cleavage of synthetic peptide substrates. In both cases their enzymatic properties were consistent with those of the native homodimeric HIV-1 PR. The kcat values for the Control, native, and covalent recombinant dimer proteases were equivalent - demonstrating that the non-genetically encoded elements of covalent structure had no adverse effect on enzymatic function. Meanwhile the observed kcat for the modified HIV-1 PR was reduced only 2-fold relative to the control PR rather than the 100-fold effect expected if both of the Ile50 NH flap-substrate hydrogen bonds contributed equally to catalysis. Apparently deletion of the flap hydrogen bond donor in one flap only has virtually no effect on the catalytic efficiency of the enzyme. The 2-fold reduction observed is consistent with only one productive substrate orientation on binding to the engineered asymmetric enzyme molecule.

The observation that hydrogen bonding is required from only one flap for full enzymatic activity in HIV-1 PR has important implications. On the basis of crystallographic observations, it had been assumed that both flaps were intimately involved in enzyme action. It was further assumed that the unique tetrahedrally coordinated water 301, and the two hydrogen bonds it mediates, played a key role in the catalytic mechanism by stabilizing the distortion of the scissile amide bond from planarity and thus increasing its susceptibility to nucleophilic attack.

Baca and Kent had themselves previously established that hydrogen bonds from the flaps to the substrate are important for the enzymatic activity of HIV-1 PR (Baca and Kent 1993). Deleting the H-bonding ability at the N(H) of Gly49-Ile50 peptide bonds of both flaps led to a 3000-fold decrease in catalytic activity. In addition to illustrating the importance of these bonds, this dramatic effect also demonstrated that other potential H-bond donors within the flaps were unable to replace the deleted interactions and restore enzymatic activity. The experiment also provided a control for their subsequent observation that H-bonding from only one flap was required for hydrolysis of a substrate Baca and Kent 2000).

However, prevailing structural data showed that the predominant binding mode for substrate derived inhibitors to HIV-1 PR involves both flaps and water 301 - even in solution. To explain this it was proposed that the coordinated binding of water 301 by both flaps forming an H-bond bridge with the substrate-enzyme complex is simply a favoured but unproductive binding mode. It was further suggested that the second flap (i.e. either flap in the native enzyme depending on the orientation of the bound substrate; or, the non-bonding flap in the modified enzyme construct) is not involved in the catalytic action of the enzyme molecule, and would remain in a highly mobile, raised, or open position - analogous to the open flaps of the RSV and SIV unliganded proteases. Structural data for the pepsin-like proteases shows inhibitor carbonyls interacting directly with a single backbone amide N(H) moiety of their single flap. Consequently, in theory, a single N(H) moiety ought to be sufficient for catalysis, even if two are available – as in HIV PR. Thus it may be that HIV PR only uses a single flap at any one time during catalysis - in a manner analogous to the pepsin-like proteases. The activity of pepsin-like proteinases additionally demonstrates that a single flap is sufficient to provide the substrate desolvation, and other features, necessary for hydrolysis. If, in the engineered enzyme, the native flap closes over the ligand in a face-on orientation (rather than the edge-on orientation seen in native enzyme crystal structures) it could supply both hydrogen bonds to the P2 and P1' carbonyls of the substrate. This would be analogous to the ligand binding mode observed in single flap pepsin-like aspartyl proteinases.

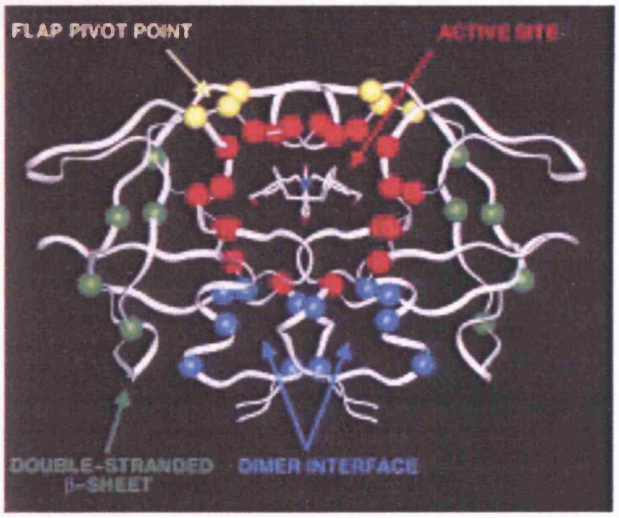
Inhibition by MVT-101 was essentially unaffected by the loss of a single flap hydrogen bond (Baca and Kent 2000) or by the loss of both flap hydrogen bonds (Baca and Kent 1993) despite the clear presence of interactions with both flaps in the crystal structure of the MVT-101 complex with native enzyme (Miller, Schneider *et al.* 1989; Miller, Geller *et al.* 1997).

Cyclic ureas, such as DMP-323 were specifically designed to replace water 301. Their sub-nanomolar potency, and their high selectivity for HIV-1 PR over cell-encoded pepsin-like proteinases, has been attributed to their ability to form two flap-inhibitor hydrogen bonds - one with each flap - via a single carbonyl group that displaces water 301 (Lam, Jadhav *et al.* 1994). It was therefore pertinent to ask whether the elimination of one of these flap hydrogen bonds would reduce their efficacy. Crystallographic analysis of an HIV-1 PR complex with a DMP323-related cyclic urea inhibitor appeared to confirm the existence of these hydrogen bonds, thus it was expected that deletion of one of them would result in a significant reduction in the binding affinity of this inhibitor. However, it transpired that binding of DMP-323 to the modified HIV-1 PR analogue is actually slightly stronger than to the Control HIV-1 PR or to the native HIV-1 PR (Baca and Kent 2000). This author conjectures that perhaps in the presence of both hydrogen bond donors the inhibitor carbonyl is centrally positioned between them by virtue of their equally attractive forces, and forms conventional hydrogen bonds with both. However, once one flap bond donor is deleted the carbonyl shifts towards the remaining hydrogen bond donor, and in achieving closer proximity a Short Strong Hydrogen Bond (SSHB) may form. SSHBs (discussed in the appendix, page 339) are much stronger than conventional hydrogen bonds, having similar characteristics to covalent bonds, and this could explain the increase in affinity that is observed. It may prove possible to determine the validity of this suggestion by means of neutron diffraction studies (also discussed on pages 81-85 this chapter, and in the appendix, pages 334-338).

Combating Resistant Viral Variants

In clinical trials HIV-1 protease inhibitors have been shown to cause a dramatic initial decline in viral plasma RNA in HIV infected individuals. However, in patients taking protease inhibitors alone, an increase in plasma viral RNA to near baseline levels occurs within one year (Kaplan 1996). Failure of HIV-1 protease inhibitor therapy is attributed to the development of resistant viral proteases.

Figure 35 Location of drug resistant mutations occurring in HIV-1 PR.



The spheres locate residues associated with drug resistant mutations of HIV-1 PR.

Figure from Ala, Huston *et al.* 1998.

The combined effects of rapid viral replication (Wei, Ghosh *et al.* 1995) and the high error rate (Williams and Loeb 1992) of HIV RT - about 1 in every 10,000 bases - result in a high frequency of viral mutation, and where these confer drug resistance they are highly favourable (Erickson and Burt 1996). Computer simulations predict that if a mutant virus has a 5% selective advantage (i.e. a replication rate 1.05 times that of the wild-type virus) it will become the dominant genotype within 100 virus generations (Erickson and Burt 1996). Consequently the virus is able to evolve under selective pressure from a protease inhibitor, becoming drug-resistant - often within a single year of monotherapy - by encoding mutants with a reduced affinity for the inhibitor but retaining the ability to process viral precursor polypeptides (Condra, Schleif *et al.* 1995; Molla, Korneyeva *et al.* 1996; Patick, Mo *et al.* 1996).

Approximately 34% of the protease residues (i.e. 34 of the 99) are associated with resistance to various inhibitors (Parikh, Larder *et al.* 2004), and a number of these are illustrated in figure 35. Mutations appear to accumulate in discrete steps initially in the active site and subsequently in peripheral or "compensatory" regions of the enzyme (Molla, Korneyeva *et al.* 1996). Mutations associated with resistance observed to date have tended to be clustered in and around the active site, reducing inhibitor binding potential (Ki values may be increased by between 100 to 1000 fold) either by limiting contacts between the inhibitor and the protease (reducing van der Waal interactions) or by competing sterically with the inhibitor (Chen, Li *et al.* 1994; Baldwin, Bhat *et al.* 1995 (3); West and Fairlie 1995; Hong, Treharne *et al.* 1996; Ala, Huston *et al.* 1997; Gulnik, Suvorov *et al.* 1995; Klabe, Bacheler, *et al.* 1998).

In addition to those associated with the active site, mutations are also seen in the flap pivot point, dimer interface, and double-stranded β-sheet (Val56-Gly78) Mutations outside of the active site probably result in conformational changes to compensate for impaired activity resulting from synchronous mutations of the active-site. These peripheral mutants may preferentially increase the affinity for native substrates over the inhibitor or possibly increase the mutant enzyme's rate of catalysis (Pazhanisamy, Stuver *et al.* 1996; Schock, Garsky and Kuo 1996). However, little is known about such compensatory mutations since their effect on the protease structure is virtually imperceptible to structural analyses (Chen, Li *et al.* 1994).

Studies have also characterised cleavage site mutants. Since mutations at the active site can affect the rate at which the Gag and Gag-Pol polyprotein precursors are cleaved by the protease, mutations at the cleavage site of these respective protein precursors could potentially compensate for reduced mutant activity. Studies show that a mutation in the P1-P6 Gag poly-protein cleavage site, in which the wild type Phe-Leu cleavage site is replaced by a Phe-Phe, renders the 184V mutant protease capable of producing a virus with 102 to 103 fold decreased sensitivity to an inhibitor, BI-LA-2011 (Erickson and Burt *et al.* 1996). However, because most drug resistant mutations occur at the active site, in theory, only a limited number of mutations can be supported by the virus since random mutations can render the protease non-functional. Thus by understanding the structural basis for resistance, new drugs can be

designed to target major classes of mutants. For example, a "structured-directed oscillation therapy" might be developed to combat resistance. This would involve treatment with a fluctuating combination of protease inhibitors designed to alternately target the mutant and wild-type enzyme and thus manipulate selection pressure to mutate the protease back and forth between the wild-type and the treatable mutant variant (Erickson and Burt 1996). Studies have shown that resistance arises because viral replication is incompletely suppressed in the plasma or tissue during periods of time between doses when drug concentrations are lowest (Molla, Korneyeva et al. 1996; Kempf, Rode *et al.* 1998). Thus inhibitors that maximise bio-availability and sustain high plasma levels are required to combat resistance.

Currently the best means of ensuring that development of drug resistance is minimised involves the use of combined therapies. In order to achieve the best blend of anti-viral agents a thorough understanding of cross resistance is required. Ideally, the drugs used in concert with each other should not have overlapping toxicity profiles and should bind to different targets. The use of multiple PR inhibitors is therefore problematic since at present they share similar chemical structures and target similar PR attributes. The combined use of PR and RT therapies has been adopted to prevent onset of drug resistance, but although multi-drug therapy has maintained the viral load in some patients at undetectable levels for 2 years, studies have reported the isolation of a replication-competent virus from their resting CD4+ T cells (Wong, Hezareh, *et al.* 1997). The existence of latently infected HIV reservoirs with the ability to mutate is already cause for concern, but it is now becoming clear that in developing drug resistance the virus may also increase in vitality (Nijhuis, Schuurman, *et al.* 1999).

It has generally been thought that the wild type HIV population is ideally fit, and that sub-optimal anti-viral therapies can therefore only result in selection of drug resistant variants with reduced fitness. Studies supporting this view have shown that suboptimal therapies select for amino acid changes conferring a selective disadvantage during replication in a drug free environment in comparison with the original wild-type virus population (Back, Nijhuis, *et al.* 1996; Goudsmit, de Ronde, *et al.* 1996; Goudsmit, de Ronde, *et al.* 1997; Eastman, Mittler, *et al.* 1998; Zhang,

Imamichi, *et al.* 1997; Zennou, Mammano, *et al.* . 1998; Mammano, Petit, *et al.* 1998).

However, the study by Nijhuis *et al.* (Nijhuis, Schuurman, *et al.* 1999), investigating the effects of ritonavir on the evolution and fitness of HIV-1 viral populations, contradicts this view. They showed that whilst initial therapy lead to the selection of drug resistant viral variants having reduced protease catalytic efficiency and replicative capacity, the continuation of that therapy eventually resulted in generation of novel protease variants that displayed no further increase in drug resistance but did exhibit increased catalytic efficiency and conferred augmented replicative ability over viruses containing the wild type protease. In their studies a quadruple mutant of the HIV-1 PR displayed a three-fold increase in activity over the patient's original wild-type PR, and this was manifested as an increase in vitality of the mutant virus over the wild-type. This study demonstrated that suboptimal antiretroviral therapy initially results in selection of drug resistant variants with impaired replication efficiencies but continued evolution in the presence of the drug allows the generation and selection of novel viral variants with increased fitness compared with the original wild-type population. This phenomenon is not limited to HIV, and increased fitness has been demonstrated *in vitro* for other RNA viruses and for bacteria (Novella, Clarke, *et al.* 1995; Novella, Duarte, *et al.* 1995; Bouma, Lenski, *et al.* 1988; Schrag, Perrot *et al.* 1997).

Phylogenetic analysis demonstrated that the mutant variants were evolved during the treatment period. Given their increased fitness over the wild-type HIV-1 it was surprising that they had not been present in the population prior to treatment. In fact at least five nucleotide changes must be selected in order to generate protease variants with an increased fitness. The initial changes must confer drug resistance, but they do so at a cost in terms of proteolytic efficiency, consequently in the absence of the drug they would not be favourable and are counter selected. Since these initial changes are a prerequisite for the subsequent mutations resulting in increased enzyme efficiency, the enzyme cannot evolve into the variant form in the absence of selection pressure from the drug. On reflection, the observations of Nijhuis *et al.* (Nijhuis, Schuurman,

et al. 1999) are in accord with Wright's concept of an adaptive fitness landscape (Wright 1932) which proposes that natural selection tends to drive a population to a local optimum, but that this is not necessarily a global optimum. Thus, a population may become trapped at a state with a suboptimal adaptation to its environment because natural selection does not allow it to pass through a trough of maladapted intermediate variants, even though a better solution may exist beyond them. It therefore appears that incomplete inhibition of viral replication creates an environment in which pre-existing drug resistant variants with an impaired replication potential are selected. Continued replication may then drive the population through a trough of maladapted variants towards a new fitness peak which may be higher than that of the original wild-type population. This may have serious implications for the design of AIDS therapies. If therapies are not sufficiently stringent they may simply result in the selection of more virulent strains than presently exist.

Alternatives to current PR inhibitors

Overcoming the problem of resistant mutants is likely to involve discovering alternatives to substrate and transition state analogue based inhibition – especially since the combination of multiple PR therapies would more effective if diverse PR attributes could be simultaneously targeted.

Anti-dimerisation and defective monomer inhibitors.

Among the possible alternatives is the design of inhibitors intended to prevent dimerisation, and thereby eliminate enzyme activity (Weber 1990; Babé, Pichuantes and Craik 1991; Zhang, Poorman *et al.* 1991; Babé, Rose *et al.* 1992; Schramm, Breipohl *et al.* 1992; Franciskovitch, Houseman *et al.* 1993; Arnold, Popinga *et al.* 1994; Schramm, Boetzel *et al.* 1996; Zutshi, Franciskovich *et al.* 1997; Caflisch, Schramm *et al.* 2000), or the use of inactive monomers to titrate out active enzyme (McPhee, Good *et al.* 1996). The dimer interface is an attractive target for therapeutic interventions (Weber 1990; Babé, Pichuantes and Craik 1991) because dimerization is essential for activity, and because it may prove possible for inhibitors to bind the PR domain within the uncleaved gag/pol polyprotein prior to particle assembly. Furthermore, although subunit-subunit interactions occur in four regions (the 'firemen's grip' at the active site; the flap region, the salt bridges, and the dimer-interface between the subunit termini) it is the four-stranded antiparallel β-sheet interface created by the C- and N-termini that provides more than 50% of the subunit

interactions (Todd, Semo and Freire 1998). The dimer interface residues are highly conserved, as is its fold, thus it may be less prone to mutation than the active site or substrate binding cleft. The strategy of using interfacing peptides to interfere with an enzyme's subunit interactions has been successfully employed to inhibit the ribonucleotide reductase from Herpes Simplex Virus (Cohen, Gaudreau *et al.* 1986 (1); Cohen, Gaudreau *et al.* 1986 (2); Dutia, Frame, *et al.* 1986), and was later developed into a potent peptidomimetic drug (Liuzzi, Deziel *et al.* 1994). The HIV-1 RT has also been successfully inhibited by means of a peptide interfering at the dimer interface (Divita, Restle, *et al.* 1994).

The use of peptide inhibitors in this way to inhibit HIV-1 PR has also been attempted. Single peptides complementary to the N or C terminal regions have been utilised as have dual stranded peptides able to interdigitate with the monomer and reproduce the typical β-sheet structure. Two octapeptides derived from the sequence of the N- and C-termini of HIV-1 protease were tested for their ability to inhibit HIV-1 reproduction, but weak inhibitory activity was found with each of the two peptides (Schramm, Nakashima, *et al.* 1991). Several other structures derived from the terminal segments have shown comparable inhibition (Schramm, Breipohl, *et al.* 1992; Babé, Rose and Craik, 1992). It has been demonstrated that the best inhibitory activity is obtained with sequences corresponding exactly to the full length β-strands in the four-stranded β-sheet region of the protease, while inhibitors derived from the C-terminus are usually better inhibitors than those from the N-terminus (Franciskovich, Houseman, *et al.* 1993).

C-terminal based peptides have been systematically modified on the basis of published PR structures, mutation studies, and sequence analysis of retroviral proteases. They were then modelled into their binding sites, and the best-fitting peptides were synthesized and tested *in vitro* (Schramm, Billich *et al.* 1993; Schramm, Boetzel *et al.* 1996). This rational design process resulted in an approximately 1000-fold improvement in terms of inhibition. These studies established a consensus sequence of potent peptidic inhibitors for positions 94-99 of the protease (Schramm, Boetzel *et al.* 1996), which can be considered as a pharmacophore for peptidomimetic or non-peptidic inhibitor design (Quéré, Wenger, *et al.* 1996; Schramm, Quéré, *et al.* 1998). Several triterpenes able to fit into the hydrophobic interface site of the relaxed monomer were identified by means of a virtual screen of the Cambridge Structural

Database using one of the pharmacophore distances, and subsequent computer docking. The use of triterpene units to build non-peptide inhibitors has several advantages, including low cost, good bioavailability, rigidity, and the provision of opportunities for side chain modifications.

Since short peptides derived from both the N and C PR termini were found to be active and had been optimised to target distinct binding sites, attempts were made to combine them into dual-sequence peptides with N and C peptide joined using linkers (Babé, Rose and Craik, 1992; Arnold, Popinga, *et al.* 1994). Other dual peptides were similarly produced by linking two N-terminal sequence based peptides (Zutshi, Franciskovitch *et al.* 1997). The cross-linkers used to generate these interfacial peptides allowed a high degree of mobility, which in turn reduced their binding efficiency. Consequently a more rigid means of linking the peptides was sought, and rigid "molecular tongs" were devised containing either tripeptidic or tetrapetidic arms linked to pyridinediol- or naphthalenediol-based scaffolds which achieved submicromolecular range PR inhibition (Bouras, Boggetto *et al.* 1999).

The use of defective monomers or non-identical PR subunits to exchange with wild type PR homodimers and produce catalytically defective heterodimers has also been investigated as an inhibitory strategy (Babé, Pichuantes and Craik, 1991; Babé, Rosé, and Craik, 1995). A structure-based approach was used to identify amino acid substitutions at the dimer interface of HIV-1 PR that would facilitate preferential association of the defective monomers. Expression of the designed PR monomers inhibited the wild-type HIV-1 PR and eliminated viral infectivity demonstrating that PR monomers can be used as macromolecular PR inhibitors. It has since been proposed that the use of defective PR monomers as trans-dominant PR inhibitors might be adopted in the clinical treatment of HIV-1 infection by gene therapy (McPhee, Good, *et al.* 1996).

The quest for active site binding, irreversible PR inhibitors resulted in the discovery that some haloperidol derivatives could also inactivate the enzyme via a sulfhydryl alkylation of the Cys95 belonging to the dimer interface (Salto, Babé, *et al.* 1994; De Voss, Sui, *et al.* 1994; Yu Caldera *et al.* 1996). Other sulfhydryl reagents such as Ellman's reagent - 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB), glutathione (Davis, Dorsey, *et al.* 1996), N-ethylmaleimide, and iodoacetamine (Meek, Dayton *et al.*

1989) were also observed to interfere with PR dimerisation in the same way. These compounds provide an alternative avenue for the design of dimerisation inhibitors.

The identification of a metal binding motif as a potential inhibitor target.

Studies involving the effects of metal binding on PR activity (Stadtman and Oliver 1991; Woon, Brinkworth and Fairlie 1992; Zhang, Reardon, *et al.* 1991) and in particular the effect of metal-cysteine interactions (Karlström and Levine 1991; Wlodawer, Miller, *et al.* 1989; Karlström, Shames and Levine 1993; Davis, Branca, *et al*, 1995; Lebon, De Rosny, *et al.* 1998; Lebon, Ledecq *et al.* 1999), and cysteine derivatisation (Davis, Dorsey, *et al.* 1996; Karlström, Shames and Levine 1993; D'Ettorre and Levine 1994; Davis, Yusa, *et al.* 1999) on PR activity have resulted in the discovery of an interesting site on the protease surface. This may provide an additional target for future drug development that is outside the realms of the active site or dimerisation region. It has been suggested that cupric ion binding to Cys67 may interfere with the correct motion of the flaps. NMR studies of HIV-1 PR have demonstrated that flap movement is accompanied by cantilever-like compensatory changes in residues 59-75 (Harte, Swaminthan, *et al.* 1990). The Cys67-Gly68-His69-Lys70 residues of this "cantilever" region have subsequently been identified as part of a metal binding sequence. Interestingly Gly-His-Lys is also the metal-coordinating motif of a copper-binding growth factor isolated from human plasma and involved in copper transportation (Karlström, Shames and Levine 1993). The binding of copper to the region, or its derivatization with DTNB, causes enzyme inactivation, and it is therefore possible that any molecule binding this region could influence enzyme activity either via an oxidation mechanism or by means of allosteric deformations.

Cysteine residues 67 and 95 as targets for intervention.

The PR cysteines (Cys67 and Cys95) are also sensitive to glutathionation, providing a possible mechanism by which the protease may be regulated through cysteine modification. It has been shown that while glutathionation of cysteine 95 abolishes activity, modification of cysteine 67 with glutathione increases activity over the unmodified protein. In addition, glutathionation at cysteine 67 markedly stabilises enzyme activity presumably by reducing autoproteolysis. The greater stability of the Cys67 glutathiolated enzyme in conjunction with an increase in activity may function to regulate polyprotein processing *in vivo* (Davis, Dorsey, *et al.* 1996; Davis, Yusa, *et al.* 1999).

Irreversible PR inhibition (suicide inhibitors).

The development of irreversible inhibitors represents another strategy to suppress the development of resistant strains. For this purpose the carboxylic groups of Asp 25 and Asp 25' represent excellent targets since their inactivation leads to complete loss of catalytic activity, and because these residues are necessarily immutable.

HIV-1 PR is inactivated by 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) (Meek, Dayton, *et al.* 1989; Rose, Rose, *et al.* 1993) and by the epoxide antifungal antibiotic cerulenin (Blumenstein, Copeland *et al.* 1989). However, to date only a few epoxide-based inhibitors of HIV-1 PR have been reported (such as UCSF 84) capable of selectively inactivating HIV-1 PR (Salto, Babé, *et al.* 1994). It has been shown that the potency of irreversible epoxide based inhibitors can be increased by extending the peptide sequence (Lee, Choy, *et al.* 1996). NMR and topochemical studies of irreversible inhibitors containing a cis-epoxide as amide isostere have also revealed their preference for adopting extended conformations similar to the β-strand in solution (Ro, Baek, *et al.* 1998). Epoxides are highly reactive species, and *in vivo* they are likely to undergo unintended reactions prior to reaching their target, something that has limited their potential as HIV-1 PR inhibitors so far. The first results chapter of this thesis (Chapter 3) investigates the preliminary development of two irreversible or "suicide" inhibitors, including an epoxide based inhibitor.

There now follows a review of the methods used during completion of the studies described in this thesis (Chapter 2: Materials and Methods). Subsequently Chapter 3 will describe the expression and purification of HIV-1 protease with the intention of using the pure enzyme to investigate the efficacy of the two in-house suicide inhibitor designs.

Chapter Two: General Materials and Methods.

Bacterial strains used.

The strains listed in the table below were obtained from New England Biolabs, and a complete list of strains available plus a key explaining their genotypes can be downloaded as a pdf file via the internet at :-

http://circuit.neb.com/neb/tech/tech_resource/restriction/ecoli/genotypes.html

Table 1 Bacterial strains used.

E.coli strain	Genotype	Use
JM109 (Yanisch-Perron, Vieira *et al.* 1985)	F′ traD36 lacI^q Δ(lacZ)M15 proA$^+$B$^+$/e14$^-$(McrA$^-$) Δ(lac-proAB) thi gyrA96 (Nal^r) endA1 hsdR17 (r_k^- m_k^+) relA1 supE44 recA1	Used for general DNA preparation, and for protein expression using non-T7 based systems. Also used to screen T7 based DNA constructs prior to expression in BL21.
MC1061 (Wertman, Wyman *et al.* 1986; Raleigh, Trimarchi *et al.* 1989)	F$^-$ araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (Str^r) hsdR2 (r_k^- m_k^+) mcrA mcrB1	Cloning and expression of HIV-1 and HIV-2 proteases, except where stipulated otherwise.
BL21 (DE3) (Studier, Rosenberg *et al.* 1990)	F$^-$ ompT (lon) hsdS$_B$ (r_B^- m_B^- an *E.coli* B strain) with DE3, a λ prophage carrying the T7 RNA polymerase gene	Expression using T7 based plasmids (eg pRset)
BL21(DE3) pLys S	As for BL21 (DE3) above, plus the pLysS plasmid (confering chloramphenicol resistance) which encodes T7 lysozyme	Expression using T7 based plasmids. The pLys S plasmid was intended to provide tighter control of expression, and made cell lysis easier.
XL1 Blue (Bullock 1987)	F′ ::Tn10 proA$^+$B$^+$ lacI^q Δ(lacZ)M15/recA1 endA1 gyrA96 (Nal^r) thi hsdR17 (r_k^- m_k^+) supE44 relA1 lac	Used to make supercompetent cells for transformation of very low levels of DNA. Used during mutagenesis of HIV-2 PR.
BL21-CodonPlus-RIL (DE3)-RIL (Stratagene)	*E.coli* B F$^-$ ompT hsdS (r_B-m_B-) dcm + Tet^r gal endA The [argU ileY leuW Cam^r]	General protein expression where codon usage is a limiting factor.
E.coli TG1	*Sup*E hsd Δ5 *thi*Δ (lac-*pro*AB)F′[*tra*D36*pro*AB+*lacI*^{q}*lac*ZΔM15]	Used to test pKu2PR during development of pKulacI^q PR vector (this thesis).

Cloning and expression vectors.

Table 2 Summary of vectors used.

Vector	Source	Promoter	Antibiotic resistance	Description
pKulacIq	Wellcome	trc	Ampicillin	In-house expression vector
pMalcRI	New England Biolabs	tac	Ampicillin	Maltose Binding Protein fusion system, for soluble expression and affinity purification.
pKK223.3	Amersham Pharmacia Biotech	tac	Ampicillin	General cloning and expression vector, used as the basis for pKulacIq.
pTrc99A	Amersham Pharmacia Biotech	trc	Ampicillin	Expression vector, commercial equivalent of pKulacIq
pET23d	Novagen	T7 / lac	Ampicillin	Expression and shuttle vector
pTrcHIS	Invitrogen	trc	Ampicillin	Histidine tagged expression vector
pRSetB	Invitrogen	T7 and M13	Ampicillin	Histidine tagged expression vector

Source of reagents and commercial enzymes.

All enzymes used for the modification, amplification or cloning of DNA during the course of this project were supplied by New England Biolabs (NEB), unless otherwise stated in the text, and were used in accordance with the suppliers instructions. All chemicals were supplied by Sigma, unless otherwise stated in the text, and Qiagen supplied all DNA miniprep and maxiprep kits. All PCR primers were obtained 'Ready Pure' from Perkin Elmer, chromogenic and other peptide substrates were obtained from Bachem.The SPA substrate was supplied by Amersham International (subsequently Amersham Pharmacia Biotech, now GE Healthcare) to the Wellcome Research Laboratories at Beckenham. T7 sequencing kits were obtained from Pharmacia Biotech (subsequently Amersham Pharmacia Biotech, now GE Healthcare). Details of any reagents not mentioned here are given in the text as they arise.

A list of suppliers' details is provided in the appendix (page 325).

Buffers, solutions, broths and media.

The solutions and media referred to in the methods sections and in the text were all made according to the protocols described in Molecular Cloning volumes 1-3 (Sambrook 1989), unless otherwise stated in the text.

Microcentrifugation and Ultracentrifugation.

Within this text microcentrifugation refers to the use of a benchtop microfuge equipped with a rotor to accept eppendorf tubes, and assumes that the maximum RCF is selected. Typically bench top microfuges generate between 13 – 16000 g at maximum rpm. Where the RCF is critical to the application it is stipulated within the text. Here the term Ultracentrifugation refers specifically to the use of a Beckman L8-M Ultracentrifuge in conjunction with an SW28 swing-out rotor and appropriate buckets. Disposable inner tubes were supplied by Beckman to fit the SW28 rotor buckets. The instrument was operated as recommended by Beckman, and specific run parameters are given in the text.

Preparation of DNA.

Small scale or 'minipreps'.

Miniprep DNA was prepared by an adaptation of the method of Feliciello and Chinali (Feliciello and Chinali 1993).

The solution of 88% isopropanol containing 200 mM Potassium Acetate was prepared in 50 ml batches by mixing 44 ml 100% isopropanol, 2 ml 5 M Potassium Acetate solution, and 4 ml ddw in a sterile 50 ml Falcon tube.

Miniprep solutions and their function.

Solution I 25 mM Tris-Cl, pH 8.0, 50 mM glucose, 10 mM EDTA

The glucose functions to maintain osmotic pressure while the Tris buffers the cells at pH 8.0. EDTA binds divalent cations in the lipid bilayer, thus weakening the cell envelope. Following cell lysis, EDTA also limits DNA degradation by binding Mg^{2+} ions that are a necessary cofactor for bacterial nucleases. Hen egg white lysozyme was added to a final concentration of 4 $mgml^{-1}$ (immediately before use) to facilitate cell lysis.

Solution II 0.2 N NaOH, 1.0% SDS

The alkaline mixture lyses the bacterial cells, while the detergent SDS dissolves the lipid components of the cell envelope and the cellular proteins. The sodium hydroxide denatures both chromosomal and plasmid DNA into single strands leaving the intertwined circular plasmid DNA intact.

Solution III 3 M sodium acetate, 3 M acetic acid

The acetic acid restores the pH to neutral, allowing the DNA strands to renature. However, the large, disrupted chromosomal strands cannot rehybridize perfectly but instead collapse into a partially hybridized aggregate. Simultaneously potassium acetate precipitates the SDS from the cell suspension, along with the associated proteins and lipids – trapping the renaturing chromosomal DNA in the SDS/lipid/protein precipitate. Only the smaller plasmid DNA, fragments of chromosomal DNA, and RNA molecules escape precipitation and remain in solution.

A bacterial culture containing the desired plasmid DNA was grown in 5 ml LB overnight at 37°C with shaking in a 25 ml sterilin tube. The bacteria were harvested by centrifugation for 10 mins at 5500 g RCF. The supernatant was discarded, and the bacterial pellet resuspended in 250 µl of solution I before transfer to a 2 ml eppendorf tube. Cell lysis was achieved by mixing the suspension with 500 µl solution II. The precipitation of unwanted cellular proteins was achieved by mixing 750 µl of solution III with the lysate. The white pellet of precipitated protein was removed following microcentrifugation (ie. 15 minutes at 13000 rpm, and room temperature, in a benchtop microfuge - unless otherwise stated) using a sterile toothpick. The clarified solution was then mixed with 700 µl of 100% isopropanol, resulting in the precipitation of any nucleic acids present, which were then collected as a pellet by microcentrifugation. The supernatant was discarded and the pellet resuspended in 250 µl of sterile ddw containing 1 µl of RNAse solution (10 µg/ml). Incubation at room temperature for 15 minutes ensured that any RNA present would be degraded. Finally the DNA was precipitated by addition of 300 µl of 88 % isopropanol containing 200 mM potassium. The DNA was collected as a pellet by microcentrifugation, and the supernatant discarded. The DNA pellet was dried (either using a 'speedivac', or by incubation at 70°C for 5 mins), then resuspended in 10 µl sterile ddw and stored frozen (-20°C) pending use.

Large scale 'midipreps'.

Here large scale is taken to mean a plasmid DNA preparation derived from 100 ml of bacterial culture - the amount of DNA produced being dependant on the copy number of the plasmid concerned (usually 5-15 $ugml^{-1}$ of culture). Qiagen midi columns were used to prepare DNA on this scale, and the manufacturers directions were followed throughout the procedure with two exceptions. Following addition of solution III the prep was centrifuged at 5500 g for only 10 mins. Following this a white precipitate layer was removed from the surface of the solution using a sterile pipette tip, then the remaining solution was poured directly onto the pre-quilibrated column through a sterile 5 ml gilson tip which had a small wad of sterile siliconised glass wool compressed within it. The wool acted as a filter, removing any loose or floating debris that would otherwise block the column. This method had the advantage of reducing the required spin time and speed. Following elution of the DNA sample from the column, the 4 ml eluant was aliquoted into 2 ml eppendorf tubes, 1ml per tube, and

the precipitation steps were scaled down accordingly and carried out using the eppendorf tubes rather than larger centrifuge tubes. This had the advantage of using a bench top bench top microfuge, made the final resupension of the dried DNA pellet much easier, and produced ready made aliquots. Aliquots prepared in this way were ideal for sequencing or cloning use. The only disadvantage was the increased number of tubes involved, which became complicated if multiple samples were being prepared. In such cases the manufacturers protocol was adhered to.

DNA precipitation methods.

Ethanol precipitation.

The volume of DNA solution to be precipitated never exceeded 500 µl. Where a larger volume was involved it was split into 500 µl aliquots. The procedure was carried out using 1.5 ml eppendorf tubes. The DNA solution was mixed with one tenth its volume of 3 M Na acetate pH 5.2. An excess of prechilled 100% ethanol was then added, such that the eppendorf tube was filled to the 1.5ml mark. The contents of the tube were mixed by inversion, then incubated at -20°C for a minimum of 10 minutes (or overnight if desired). The precipitated DNA was pelleted by microfugation, the supernatent discarded and the eppendorf tube then spun briefly to drain any remaining supernatent to the bottom of the tube. The remaining supernatent was carefully removed using a longform glass pipette (previously extruded in a bunsen flame to provide a fine tip). The pellet was then allowed to air dry, or was dried either by 'speedivac' or in an oven at 70°C for 5 minutes.

Precipitation at -20°C gave a better yield than incubation on ice. Yield was also proportional to the microcentrifugation time, and 15 minutes was considered a minimum. Also note that for all such precipitation steps the eppendorf tube was located in the microfuge rotor such that the lid hinge was outer-most. This ensured that the position of the DNA pellet after microcentrifugation could always be predicted.

Isopropanol precipitation.

This method was used both as an integral part of the miniprep method of Feliciello and Chinali (Feliciello and Chinali 1993) (See: Preparation of DNA; Small scale or 'minipreps'), and subsequently as a general replacement for the ethanol precipitation step above. It was used for the general reduction in volume of DNA solutions, removal of enzyme and buffers in between double restriction digests, and precipitation of DNA samples prior to loading onto agarose gels. It was carried out at room temperature (which was optimal), there was no need to incubate following addition of the isopropanol solution (thus it was quick), and it proved consistently effective. The 88% isopropanol / 200 mM Potassium Acetate solution was prepared as previously described (See: Preparation of DNA; Small scale or 'minipreps').

The volume of DNA solution to be precipitated was adjusted to precisely 250 µl, either by addition of ddw or by splitting it into appropriate aliquots. The DNA solution was then mixed by inversion with exactly 300 µl of the isopropanol / Potassium Acetate solution. The precipitated DNA was collected as a pellet by microcentrifugation, and the supernatent discarded. In order to remove all traces of the supernatent the eppendorf was briefly spun again (approximately 30 seconds) and any remaining supernatent was sucked out using a fine tipped glass pipette. The pellet was then allowed to air dry, or was dried either by 'speedivac' or in an oven at 70°C for 5 minutes. The DNA was then resuspended as required in either ddw or TE buffer, and stored or used as desired.

Polymerase Chain Reactions

A variety of PCR based techniques were used for purposes including basic DNA amplification for cloning or as a diagnostic tool, site directed mutagenesis, whole gene synthesis, and ABI sequencing.

In principle the PCR method described by White *et al.*, and Erlich respectively (Erlich 1989; White, Arnheim *et al.* 1989) was followed for all applications. However, for mutagenesis Stratagene's QuikChange^{Tm} method was employed – see Site directed mutagenesis, below.

For gene synthesis the Recusive PCR method of Prodromou and Pearl (Prodromou and Pearl 1992; Wheeler, Prodromou *et al.* 1996) would have been used (see Chapter 6 Results and Discussion). ABI sequencing reactions were carried out using an ABI Prism^{TM} kit (Applied Biosystems), and following the manufacturers instructions.

Choice of polymerase

Taq polymerase was used for general PCR reactions since it was reliable and cheap. Vent and Pfu were used when higher fidelity was required (eg during gene synthesis or site directed mutagenesis).

Additional components

Reactions occasionally required the presence of additional components in the reaction mixture to achieve successful amplification. The component required depended on the polymerase used - $MgCl_2$ (at 0, 2 and 4 mM) was used for Taq, $MgSO_4$ (0, 2 and 4 mM) for Vent, and DMSO (0 and 5 %) for Pfu.

Primer design

General PCR primers were designed manually, according to the Wallace rule (Itakura, Rossi *et al.* 1984), with a view to obtaining an optimum annealing temperature (or melting temperature T_m) and incorporating the necessary restriction sites for the task in hand. Typically each primer had a 5′ GC rich non-coding clamp region, followed by the desired restriction site sequences, and subsequently a stretch of coding sequence, designed to anneal to the template DNA, ending in a GC rich coding region intended to act as a 3′ GC clamp. The 5′, GC rich, non-coding region, besides acting as a clamp, also served to flank the restriction enzyme sites. This enables some restriction enzymes to cut their respective site more efficiently.

Primers used for site directed mutagenesis were designed as complementary pairs, with the mutant bases located centrally, flanked each side by a stretch of coding sequence. This is described in more detail below (Site directed mutagenesis).

Primers for gene synthesis by recursive PCR were designed with the aid of computer programs written especially for that purpose, since the design parameters were too complex for this task to be successfully achieved manually. This process is described in Chapter 6.

Site Directed Mutagenesis.

The QuickChange site directed mutagenesis kit (supplied by Stratagene) was used according to the manufacturers instructions. Information can be found at http://stratagene.com/homepage/, and a pdf file of the QuickChange manual can be downloaded from http://stratagene.com/lit/manuals.aspx. Briefly, a complementary pair of primers was designed for each mutation reaction, both containing the desired mutation in the centre of the primer. Such primers must be between 25 and 45 bps long, with10-15 bases of correct sequence each side of the mutation site, and having a $T_m \geq 78°C$ as determined by the formula:

$T_m = 81.5 + 0.41\ (\%GC) - 675/N - \%$ mismatch

Where N = primer length, and values for %GC and % mismatch are whole numbers.

Optimally the primers should have a GC content of 40% and should terminate in a clamp of one or more G / C bases. The primers need not be 5′ phosphorylated, but must be HPLC or PAGE purified since impurities decrease mutation efficiency. In the final reaction mixture the primer concentration must be in excess over the template.

Reaction mixture typically consisted of 5 µl 10 x reaction buffer, 5 – 50 ng of dsDNA template, 125 ng of each primer, 1 µl of dNTP mix, and ddw up to a final volume of 50 µl. The enzyme, 1 µl of *Pfu* Turbo DNA polymerase at 2.5 U/µl, was added last. The reaction cycles consisted of 30 seconds at 95°C, then 12-16 rounds of; 95°C 30 s, 55°C 60 s, 68°C 120 s per kb of template plasmid length. 12 rounds were used for point mutations, 16 for single amino acid changes. Following this the reaction was cooled on ice for 2 minutes. The non-mutated supercoiled parent dsDNA was digested following the addition of 1 µl of *Dpn*I restriction enzyme to the mix and incubating at 37°C for 1 hour. Following this digestion step the remaining mutated DNA was

annealed by cooling on ice, then used to transform XL1-Blue supercompetent cells (see Chapter 2, Transformations, Preparation of supercompetent cells). Supercompetent cells were used to increase the probability of success, especially since the PCR reaction only incorporated up to 16 cycles (limiting the product yield). Transformants were picked, grown and their DNA miniprepped for diagnostic restriction enzyme digest (where incorporation of a site was possible simultaneously with the incorporation of a mutation), and sequencing.

Sequencing of DNA

Conventional double stranded DNA sequencing

Sequencing was carried out by the method of Sanger *et al.*, using a T7 kit from Pharmacia Biotech, and largely in accordance with the supplied instructions, but with the modifications described below.

Double stranded DNA to be sequenced was prepared using Qiagen midiprep kits (previously described). Pharmacia Biotech supplied the sequencing primers (T7 forward and reverse), and where these were not appropriate PCR primers for the clone of interest were used instead.

Template DNA (250 µl or approximately 25 µg) was isopropanol precipitated, then resuspended in 10 µl ddw ready for denaturing and primer annealing. The denaturing and annealing procedure was carried out in one PCR style step, obviating any need for the alkaline denaturation, neutralisation and precipitation steps of standard protocols. The 10 µl of DNA template was mixed with 2 µl of primer and 2 µl of annealing buffer (as supplied in the T7 kit) in a small eppendorf. Primer concentrations were previously adjusted to between 2.5 and 5 µM (so that 2 µl contained 5-10 pmol). The mixture was then incubated on a PCR block programmed as shown in table 3.

Table 3 PCR amplification program.

94°C	10 mins	(heat denaturing step)
65°C	5 minutes	(gradual annealing steps)
37°C	10 minutes	
25°C	10 minutes	
4°C	∞	(followed by storage on ice until ready to use)

The $^{35}S\alpha$ dATP was supplied in 100 μCi aliquots from a centralised stock at the Biochemistry department UCL. This reagent was never more than 1 month old. Preparation of the sequenase enzyme, the labelling reactions and the other reagents was carried out in accordance with the manufacturers supplied protocol. However, the reaction, incubation, and heating steps were carried out using a 48 well microtitre plate rather than individual eppendorf tubes. The first four columns across the top of a plate were labelled A C G T, and the rows down the side were labelled with an appropriate sequence name. Reactions were then set up in the appropriate wells. This allowed easy manipulation, increased speed, ease of storage, and reduced the risk of confusion between reactions that could occur using multiple tubes. Each labelling reaction was first dispensed into the bottom of its respective well, then the appropriate sequencing mix was dispensed onto the side wall of the same well. Once the required time had elapsed (as stipulated by the protocol) the sequencing mixes were simultaneously added to the labelling reactions by simply tapping the plate sharply on the bench to dislodge the droplets on the sides of the wells. Mixing was achieved by rapid swirling of the plate, and the plate was incubated at 37°C by floating it in a water bath. As the incubation step neared completion the plate was removed from the bath. Stop mix was dispensed onto the side of each well, and at the desired moment a sharp tap was again administered so that the stop mix droplets were dislodged stopping all reactions simultaneously.

The complete plate of reactions was covered using Nesco film, and either floated on a hot water bath (80°C) for 2 minutes ready for loading onto a gel, or stored at -20°C pending analysis.

Conventional sequencing gels

The stock solution for preparation of sequencing gels is shown below (table 4). 'Ultrapure Protogel' acrylamide was used (30% acrylamide, 0.8% bis-acrylamide) and was purchased ready made from National Diagnostics. Both these solutions were stored in the dark at 4°C.

Table 4 6% Stock solution for sequencing gels.

250 g urea dissolved in 125 ml ddw
plus 50 ml TBE (10x stock)
plus 100 ml 30% acrylamide / bis stock.
Solution microwaved on medium setting for 10 seconds to dissolve urea, then made up to 500 ml with ddw and 0.45 μm filtered.

The gel mix was prepared immediately before pouring, using 75 ml of the above stock plus 750 μl of fresh 10% APS solution, and 32.5 μl TEMED. The constituents were added, mixed using a vortex, and used immediately.

A vertical slab apparatus (Raven) was used to run the gels. The gel was poured between two glass plates (50 cm x 20 cm) separated by a spacer down each edge (0.4 mm). The casting method as described by Sambrook *et al.* (1989) was modified as follows. Unconventionally, the plates were not taped up at the sides and bottom, but were instead held together by bulldog clips (from Niceday) - two on each side near the bottom and middle of the plates. The plates were then positioned horizontally on the bench, raised slightly at the top end by resting them on a Duran bottle lid. The gel mix (described above) was prepared, and quickly drawn into a 50 ml sterile syringe. The mix was gently applied by syringe between the plates from the top end, using a sweeping side-to-side motion to ensure even distribution across the width of the plates. Simultaneously the plates were gently tapped at the leading edge of the gel solution with a pipette handle to help maintain an even flow and prevent bubble formation. With this method capillary action drew the gel mix down the plates until they were completely filled. Then the Duran lid was removed and the plates layed flat, stopping further movement of the gel solution. The combs were then positioned flat side first (ie upside down) into the top of the gel, between the plates, and held tightly in place with two further bulldog clips. The gel was then left to set.

This method was faster and simpler than conventional methods of the time, rarely incorporated bubbles, never leaked, and did not require messy tape. Leaking was common with the conventional method and (ironically) was in fact a product of pouring vertically while using tape.

In all other respects gels were set up and run according to the method of Sambrook *et al.* (1989), and as recommended by the manufacturers of the sequencing apparatus used. Gels were pre-run at 55 W for 30 minutes to warm them to running temperature, and wells were flushed clear of urea prior to loading using a plastic Pasteur pipette. Sequencing reactions were incubated at 80°C for 2 minutes prior to loading, and 2 µl of reaction mix was loaded per well using a Gilson P10 pipette. Electrophoresis was carried out at a constant voltage of 37 W until the lower dye front reached the bottom of the gel. Occasionally a duplicate batch of reactions was loaded at this point, and electrophoresis recommenced until the lower dye-front from the second loading reached the bottom. In this way the range of readable sequence was increased.

On completion of electrophoresis the gel plates were carefully prized apart, ensuring that the gel stuck to one plate only. The gel was then removed from the plate by lowering an appropriately sized square of dry Whatmann 3MM paper evenly onto the gel, and pressing firmly. The paper was then gently lifted from one end, and the gel was removed from the plate having stuck to the paper. The conventional fixing step was omitted since it was found to be unnecessary, and made the transfer of the gel from plate to paper more problematic. This modification also shortened the drying time required (30 minutes at 80°C under vacuum, using a BioRad slab gel dryer).

Fuji medical X-ray film was exposed to the dry gel overnight (or longer if required), and subsequently processed using an autoradiograph developer (from AGFA).

Fluorescent Automated DNA Sequencing

Automated DNA sequencing was carried out using the ABI Prism system from Applied Biosystems, and the protocol supplied by the manufacturer (details available at www.appliedbiosystems.com). Completed samples were sent for analysis to Cambridge Biosciences.

DNA manipulations

Agarose gel electrophoresis

A 0.5-1% w/v agarose solution was prepared in either TBE or TAE (buffers prepared as described by Sambrook 1989) and the melted agarose dissolved into the buffer by microwave heating. Ethidium bromide was added to a final concentration of 0.5 $mgml^{-1}$, and the solution allowed to cool until approximately 55°C before pouring. The gel tray to be used was prepared according to the relevant manufacturer's instructions (i.e. with damns in place, and combs positioned with the teeth 1-2 mm above the tray surface). The gel melt was poured to a depth of approximately 5 mm and allowed to solidify. The gel was placed in the running chamber, immersed in the appropriate running buffer (1 X TBE or TAE), also containing ethidium bromide at 0.5 $mgml^{-1}$, and the comb was removed to form sample wells. Samples (including size standards) were prepared as described below, loaded into the wells (approximately 10 µl per well), and the chamber lid fitted. Electrophoresis was carried out using 50-100 volts until the dye markers had migrated sufficiently (dependant on size of DNA analysed). After electrophoresis the gel was observed over a UV light box (set at short wave for visualization, or long wave for band extraction – since long wave UV causes less DNA damage). UV absorbant goggles and face shield were used throughout the visualization process. Photographs were taken as required.

Samples were prepared by adding 1µl of 10 X DNA loading buffer per 10 µl final volume of sample solution. Size standards were either of bacteriophage λ DNA cut to completion with PstI restriction endonuclease, or commercially produced DNA ladders supplied by NEB. DNA loading buffer consisted of 0.21% bromophenol blue, 0.21% xylene cyanol FF, 0.2 M EDTA, pH 8.0 and 50% glycerol in sterile ddw.

Restriction enzyme digests

Restriction enzyme digests were typically completed in a final volume of 100 µl, comprising 10 µl of the appropriate 10x reaction buffer (supplied with the enzyme used), usually 1-2 µl of enzyme (but never more than 10 % of the final reaction volume)[7], and the remainder DNA solution in water. Digests were carried out in sterile 500 µl eppendorf tubes, at 37°C (unless otherwise stipulated by the supplier of the enzyme used) overnight. Double digests, where DNA was incubated with two

[7] because enzymes are stored in glycerol, and greater than 10% glycerol in the reaction can be inhibitory.

enzymes simultaneously, were carried out using a buffer / temperature suitable for both enzymes (as advised by the supplier). Where enzymes were incompatible, separate digests were carried out interspersed by a clean up step or an ethanol (or isopropanol) precipitation. Analysis of a small aliquot (usually 10 µl) by agarose gel electrophoresis was used to assess the success of a digest.

Band purification of DNA

Extraction of DNA bands from agarose gel was used to isolate a desired fragment from other contaminating fragments (eg. during the excission of a gene from an unwanted plasmid). Separation of the DNA fragments was achieved using the standard agarose gel electrophoresis procedure (though it is essential to use freshly prepared agarose), the gel was then viewed under long wave ultraviolet light (which is less deleterious to the DNA than shortwave UV) and the desired band excised using a clean scalpel. A number of methods could be used to extract the DNA from the excised gel, the most reliable involved using the QIAquick gel extraction kit from Qiagen (used according to the manufacturers instructions).

Dephosphorylation.

This was carried out using Calf Intestinal Alkaline Phosphatase (CIAP or CIP), which catalyses the removal of 5' phosphate groups from DNA, RNA and ribo and deoxyribonuclease triphosphates. Since CIP treated fragments lack the 5' phosphoryl termini required by ligases, they cannot self ligate (Sambrook 1989). This property has been used here to decrease the religated vector background seen during some cloning strategies. Following the digestion by restriction enzyme of the DNA in question, any remaining enzyme was removed using 'Strataclean Resin' as recomended by the manufacturer (Stratagene). Following this the DNA was precipitated using either isopropanol or ethanol (as previously described) and the pellet resuspended in 100 µl ddw and 10 µl CIP x 10 buffer (as supplied). To this solution was added 1µl of CIP enzyme, and the resultant reaction mixture was incubated at 37°C for 30 mins. Next 5 µl of 500 mM EDTA pH 8.0 was added and the mix incubated for 10 mins at 75°C. The combined effect of heat and the chelation of Mg from the buffer served to inactivate the CIP enzyme. Subsequently a further three Strataclean extractions were carried out (20 µl of slurry per extraction). Finally the DNA was precipitated by addition of $1/10^{th}$ volume of Na acetate pH 7.0 (Not pH 5.2, otherwise EDTA would not remain in solution), and ethanol followed by

incubation at 20°C for 5 mins and centrifugation at 13000 rpm for 15 mins (a chilled 70% ethanol wash step is optional). The DNA pellet was dried using a speedivac, then resuspended in ddw as required.

Generating blunt ended DNA using Klenow and T4 DNA polymerase enzymes.
Where the DNA insert and the vector MCS contained incompatible sites blunt ended ligation was utilised. The insert DNA was first cut with an appropriate restriction enzyme to yield an overhanging end, then the overhang was converted to a blunt-end for ligation to the vector. Both Klenow (DNA Polymrase I Large Fragment) and T4 DNA Polymerase were use (as appropriate) to fill 5′ protruding ends with dNTPs since both enzymes have 5′ to 3′ polymerase activity. T4 DNA Polymerase was also used to polish 3′ protruding ends in the presence of dNTPs by exploiting its robust 3′ to 5′ exonuclese activity The enzymes used were supplied by Promega, and the protocols used are described in the Promega Corporation "Protocols and Applications Guide" (Third Edition - 1996), and the Promega Corporation "Enzyme Resource Guide: Cloning Enzymes" (#BR075B).

Strataclean resin
'Strataclean' resin was used to extract unwanted restriction enzymes, and CIP, from DNA solutions where they may interfere with subsequent reactions, such as ligations. The resin was used as supplied by Stratagene according to the following method. The volume of resin used per extraction was equal to the square root of the volume of DNA solution to be cleaned. Certain enzymes require multiple extractions in order to ensure that they are completely removed from the solution (this information is available from Stratagene, and is supplied with the resin). The appropriate volume of resin was added to the DNA solution and the mixture vortexed for 15 seconds, and incubated at room temperature for 1 minute. The resin was pelleted by centrifugation (1 minute at 13000 rpm, benchtop microfuge), and the supernatent removed to a fresh tube. Care was taken to avoid carry-over of any resin to the fresh tube, and only a volume corresponding to the original volume of the DNA solution was taken. Any remainder was discarded along with the unwanted pellet. A second spin further ensures that there is no carry-over. As an alternative 0.2 μm microfuge spin filters (available from Sartorius) were used to filter out any remaining resin.

The slurry bound unwanted protein, removing it from solution. Subsequently the resin pellet and bound protein were discarded while the DNA remained in solution and was retained with the supernatent. The resin represented a useful alternative to phenol chloroform, especially where small (µl) volumes were involved, and sample losses would otherwise have been high. However, this process also involved a small but inevitable loss of some DNA.

Ligations

DNA ligations were carried out typically in a final volume of 20 µl. Vector and insert DNA were mixed in a ratio of 9 µl : 9 µl (given that their concentrations were approximately the same), melted by incubation at 50°C for 2 - 5 minutes, then left to cool slowly to room temperature (allowing annealing to occur). Alternatively a process identical to that used for the annealing of sequencing primers to double stranded template using a PCR block (see Conventional double stranded DNA sequencing) was used. Following the annealing process, the ligation mix was stored on ice, and an appropriate volume (here 2 µl) of x10 ligation buffer was added to the cooled DNA mixture, followed by 1 µl of T4 DNA ligase (as supplied by NEB). The mixture was then incubated over night either at room temperature or in a cold room water bath set at 16°C – reputed to be the optimum temperature for ligation reactions (Sambrook 1989).

Transformations

Here the term 'transformation' (first used by Fred Griffith in 1928) indicates the uptake of plasmid DNA by bacterial cells (though transfection would perhaps be more accurate and avoid possible confusion with the use of 'transformation' in relation to the process of tumorigenesis). The state of a bacterial cell in which it is able to take up plasmid DNA (ie. be 'transformed') is termed 'competence'. It has been observed that bacterial cells treated with ice cold $CaCl_2$ become competent, and when exposed to a brief heat shock can be transformed with any plasmid DNA present in their suspension (Mandel and Higa 1970; Cohen, Chang *et al.* 1972; Higa and Mandel 1972; Mandel and Higa 1992). This is the basis of all transformation methods (with the exception of electroporation).

The preparation of competent cells.

Transformation efficiency for a given plasmid is defined as the number of transformants per μg of DNA (Hanahan 1983), and the method described here typically yielded cells with an efficiency not less than 10^7 transformants per μg, and up to 10^9 transformants per μg (results not shown) using the plasmid pRsetB. These cells were sufficient for most transformation purposes. On occassions where the concentration of DNA was very low, for example during procedures such as mutagenesis, supercompetent cells (efficiency of up to 10^{11} transformants per μg) were used. The production of supercompetent cells is described later in this section.

A 5 ml volume of LB was inoculated with the desired *E. coli* strain, either from freshly grown colonies on an agar plate or from a glycerol stock. Appropriate antibiotic was added if required (eg. in the case of BL21 (DE3) plys S, chloramphenicol would be used), and the culture incubated with shaking at 37°C overnight.

The overnight culture was used to inoculate 200 ml of prewarmed LB, which was incubated with shaking at 37°C, the OD_{600} (Optical Density measurement at a wavelength of 600 nm, made using a spectrophotometer and plastic 1 ml cuvette) of the culture being recorded at regular intervals after the first $1^1/_2$ hours in order to assess the correct stage for harvesting. Studies of cell competency vs OD_{600} (Tang,

Nakata *et al.* 1994) have shown that there are two peaks of optimal competency during the growth of an *E.coli* cell culture, the first occurs at approximately $0D_{600}$ = 0.3 - 0.4 and the second at OD_{600} = 0.94 - 0.95. The latter of these two peaks provides cells of the greatest competency, however, the peak tails off sharply and can easily be missed. The earlier peak provides slightly less competent cells, but has the advantage that the peak is broad and in any case if it is missed the culture can be grown to reach the second peak. Between these peaks there is a trough in the levels of competency with a minima at $0D_{600}$ = 0.7. Typically cells were harvested on reaching an OD_{600} = 0.94 consistent with the second peak of competency.

Harvesting was carried out by transferring the culture to four 50 ml Falcon tubes, followed by centrifugation at an RCF of approximately 5500 g for 5 minutes at 4°C. The supernatent broth was poured off, and each pellet resupended in 5 ml of ice cold 100 mM $MgCl_2$ prior to being made up to 12 ml with the same solution. The centrifugation step (above) was repeated and the supernatent $MgCl_2$ solution discarded. The pellets were then resuspended in 5ml of ice cold 100 mM $CaCl_2$ prior to being made up to 25 mls with the same $CaCl_2$ solution and incubated on ice for between 30 and 90 minutes. Following this the centrifugation step was repeated, the $CaCl_2$ supernatent was discarded, and the pellets very gently resuspended in 2 ml of 85 mM $CaCl_2$ solution containing 15% glycerol. The total 8 ml of competent cell suspension was finally dispensed into prechilled sterile eppendorfs in 200 µl aliquots, then frozen at -70°C for storage. Aliquots were thawed on ice immediately prior to use, but it should be noted that thawed and refrozen cells ceased to be competent.

Preparation of supercompetent cells

Supercompetent cells (ie. competent cells yielding 10^{11} transformants per microgram of DNA) were used for transformations where DNA concentrations were low (eg. after a mutagenesis reaction), and the chances of obtaining colonies with ordinary competent cells would therefore have been remote. Typically XL1-Blue cells were used for this purpose, although JM109 cell could also be used if required.

Table 5 Media and Buffers required to prepare supercompetent cells.

SOB 20 g Tryptone, 5 g Yeast Extract and 0.5 g NaCl were dissolved in 1 litre of ddw and the solution was autoclaved. The following solutions were made separately, sterilised by 0.2 μm filtration and added in the amounts stipulated after the media had cooled: 10 ml 250 mM KCl, 5 ml 2M $MgSO_4$ and 100 μl 5M NaOH.
TB 100 ml of 10 mM PIPES buffer containing 15 mM $CaCl_2$ and 250 mM KCl was made, and adjusted to pH 6.7 with KOH. $MnCl_2$ was added to a final concentration of 55 mM, and the solution sterilised by 0.2 μm filtration.

A 100 ml volume of SOB medium (table 5) contained within a sterile 500 ml flask was innoculated with the desired bacterial strain and incubated with vigorous shaking at 18°C until an OD_{600} of 0.6 was reached. This typically took in excess of 30 hours. At this point the culture was chilled on ice for 10 minutes, then harvested by centrifugation at a low RCF (less than 2000 g) using a chilled rotor. The supernatent was carefully removed, and the cells were gently resuspended in 30 ml of TB (table 5). The resuspension was left to stand on ice for 10 minutes, then centrifuged as before. The supernatent was again removed, and the cells resuspended in 8 ml of TB containing 7% DMSO then left to stand on ice a further 10 minutes. The cell suspension was aliquoted into pre-chilled eppendorf tubes (typically in 200 μl portions), and each aliquot was flash frozen using liquid Nitrogen. Aliquots were stored at -70°C, and thawed on ice immediately prior to use. As with normal competent cells, thawed and refrozen cells ceased to be competent.

The transformation procedure.

An aliquot of frozen competent cells was thawed on ice, then gently resuspended by lightly flicking the eppendorf tube. The DNA to be used for the transformation was added to the cell suspension and the mixture incubated on ice for between 30 - 90 minutes. The eppendorf was then transferred to a heat block or water bath preset at 42°C and incubated for 3 minutes (or 37°C for 5 minutes). Immediately following this incubation the tube was returned to ice and left a further 5 minutes. The cells were then spread onto dried warm LB agar plates containing the appropriate antibiotic(s) (50 - 75 µl cell suspension per plate), and the plates incubated at 37°C overnight. It was found that the addition of warm LB to a cell suspension followed by incubation for 1 hour at 37°C prior to plating out (traditionally recomended as a means of overcoming 'phenotypic lag', allowing expression and build up of the phenotypic properties conferred by the plasmid before exposure to antibiotic selection) was counter productive. Experiments (results not shown) confirmed that more transformants survived when cells were immediately spread onto plates. This was true for all transformations carried out during the work described in this thesis, regardless of strain or resistance marker used. It also had the added advantage of saving time, and was adopted as a standard procedure. It can be explained at least in terms of ampicillin resistance, since there is no phenotypic lag with this marker. Resistance develops rapidly, and ampicillin effects cell wall biosynthesis only in cells that have progressed into active growth (Old and Primrose).

Testing of competent cells.

Transformations are often the final stage of cloning procedures involving many complex and time-consuming stages. Indeed, the DNA used in a transformation sometimes represents a unique sample - making it very important to ensure that each new batch of cells is actually competent. Tests were carried out on each new batch of competent cells prior to their use. All cells were tested after freezing to ensure that the freezing process itself had not caused any deleterious effect. A single 200 µl aliquot, thawed on ice, was normally sufficient to carry out all three tests.

Table 6 Competent cell tests.

• **Vitality**	Cells were still alive.
• **Resistance**	Cells were not antibiotic resistant (unless intended to be).
• **Competence**	Cells were competent

Vitality.

50 µl of thawed cell suspension was spread aseptically onto a fresh LB agar plates devoid of any antibiotic. The plate was then incubated overnight at 37°C. Confluent cell growth was expected. No cell growth indicated that the cells had not survived the process, and the batch was discarded.

Resistance.

50 µl of thawed cell suspension was spread onto each of several LB agar plates containing different antibiotics. The plates were incubated at 37°C overnight. No cell growth on agar plates containing antibiotic was expected, unless those cells were already known to contain a resistance gene. Where cell growth did occur it indicated the presence of unwanted resistance and the whole batch was discarded to avoid a high background during transformations. Cells were routinely tested for resistance to ampicillin and chloramphenicol. If the use of other antibiotics was anticipated then cells were tested for resistance to these also.

Competence.

50 µl of thawed cells, transformed with approximately 1 ng of a control plasmid (usually a cloning or expression vector), was spread onto an LB agar plate containing the appropriate antibiotic. The plate was incubated overnight at 37°C, and the extent of cell growth indicated the level of competency. This test required that the cells were also shown to have no prior resistance to the antibiotic used (see 'Resistance' above).

Screening clones for the presence of an inserted gene or DNA fragment

Quick screen

When a new gene or PCR product is ligated into a cloning or expression vector it is necessary to select the correct construct from a background of incorrect constructs and empty vectors. Potential clones can be selected after a ligation by comparing the relative sizes of native plasmid DNA and transformant DNA. If the plasmid DNA isolated from a transformant is the same size as the native plasmid then it is unlikely to contain the insert of interest. However, if it is larger by an amount corresponding to the size of insert, then it is likely to contain that insert. A 'Quick Screen' allows this selection process to be carried out very quickly, and on a large number of samples simultaneously.

Materials required include the Quick Screen Buffer (QSB), shown in table 7, and sterilised toothpicks.

Table 7 Quick Screen Buffer.

1 X TBE buffer containing 6% glycerol, 1% SDS, 10 $\mu g ml^{-1}$ RNAse, and 0.002% Bromophenol blue.

A small amount of cells from the colony to be screened are scraped up with a sterile toothpick, and then transferred to an eppendorf containing the Quick Screen buffer. The toothpick is swiftly twisted in the buffer to release the cells into the buffer, and mix them. The buffer causes cell lysis to occur (this happens in a few minutes, usually while other samples are being prepared for the same screen), releasing plasmid DNA into the solution. The tubes are then spun in a bench top microfuge for 10 minutes. The solutions can then be directly loaded onto an agarose gel for electrophoresis. A gelatinous pellet forms during centrifugation, and this can be removed from the tubes by pipette. Only the clarified supernatent should be loaded onto the gel. Cells containing the control plasmid (with no insert) should be prepared in the same way, and loaded at either end of the gel to provide a size standard. Once the gel has been electrophoresed, the ethidium bromide stained plasmid DNA can be viewed using a UV transilluminator. Lanes containing plasmid that is larger than the control (ie. the

bands are higher up the gel) correspond to transformants that are worth further investigation.

Restriction digest screen

Potential clones are selected from unwanted background, or mis-incorporated insertions, by preparing their DNA by miniprep and incubating it with restriction enzymes. The enzymes are selected such that they will release the inserted fragment or gene from the vector, if it is present. Usually the same enzymes used to create the construct are also used to digest it. The presence or absence of an insert is revealed following digestion by agarose gel electrophoresis. This screen has the advantage of also demonstrating that the restriction sites used to make the construct are still recognisable by the enzymes - in other words the junctions between vector and insert are likely to be correct. It is a lengthy process, however, involving growth of cells, minipreps of their DNA, and incubation with (possibly expensive) enzymes. Frequently this was used as an additional screen used on clones, which had proved positive with the Quick Screen (above). Transformants were picked from the original plate and streaked onto a fresh one using a toothpick. The same toothpick was then dropped into a sterilin tube containing 5 ml of LB (supplemented with the appropriate antibiotic). Both the fresh plate and the 5 ml culture were grown overnight at 37°C. The streaked cells were then used for the Quick Screen, and pellets from cultures corresponding to Quick Screen positives were then used to provide miniprep DNA for restriction screens and further transformations.

PCR screen

The presence of a gene of interest in a new construct can be demonstrated using a PCR screen. This is a conventional PCR using the construct under investigation as the template, and primers, which will amplify the target gene if it is present. The completed reaction is analysed using an agarose gel for the presence of amplified product (indicating that the desired gene is present). A construct, which does not contain the target gene, is used as the template for a control reaction to ensure that there is no non-specific amplification. Usually miniprep DNA was used as the templates, but it is also possible to use lysed bacterial cells directly.

Protein expression screens

The presence of a gene of interest was also demonstrated by detection of the desired protein following induction of protein expression. This had the advantage of allowing selection of clones able to express the required protein. However, it had the disadvantage of not allowing detection of genes which are present but which for some reason (eg toxicity or poor codon usage) are not overexpressed. It also could not necessarily distinguish between a construct which was correct and a construct, which was incorrect but produced a protein of the expected size. Protein expression was therefore used to select clones in conjunction with DNA based screens, rather than in isolation.

Small scale (5 ml) cultures were prepared as described in the next section (see: Bacterial cell culture for protein expression - Small scale protein expression) and the uninduced cells compared with induced cells by means of SDS PAGE analysis, Western blots and Dot blots (as described in the section: 'Protein preparation and analysis', page 153). Activity assays were used where the expressed protein was expected to exhibit an activity, which could be assayed in some way (assays used are described in: 'Protein preparation and analysis' page 163).

Bacterial cell culture for protein expression.

Small scale protein expression

In order to maximise protein expression, small-scale trials using 5 ml culture volumes were undertaken. Generally a colony was used to innoculate 5 ml of LB, containing sufficient antibiotics to maintain selection of the plasmid. This was grown at 37°C overnight and 50 µl then used to inoculate a fresh 5 ml broth. Once the fresh culture reached mid-log phase, as determined by an OD_{600} = 0.5, 500 µl of the broth was removed for analysis (uninduced sample), and the remaining 4.5 ml was induced by addition of IPTG to a final concentration of 1 mM. The culture was grown for a further 4 - 6 hours, during which time the OD_{600} was measured at hourly intervals and 500 µl aliquots of cell culture removed at each interval for analysis. The cell aliquots were harvested by microcentrifugation in eppendorf tubes, and the cell pellets were resuspended in distilled water.

The volume of water used to resuspend the cell pellet harvested from 1 ml of culture was equal to OD_{600} value for that culture divided by 4. Thus the cell pellet from a 500 µl culture of OD_{600} = 0.6 would be resuspended in a volume of 100 µl (see working below, and table 8).

$$(0.6 \div 4) \times 0.5 \text{ ml} = 0.1 \text{ ml}$$

Table 8 Standardised resuspension of cells for SDS-PAGE analysis.

$(OD_{600} \div 4)$ x culture volume (ml) = resuspension volume (ml)

By preparing each cell pellet in this way their cell densities were standardised, allowing a direct comparison of protein expression levels to be made between aliquots. This method also ensured that provided 15 µl of cell suspension was loaded per lane the gel would not be overloaded.

For whole cell protein analysis 15 µl of cell resuspension was added to 5 µl of SDS PAGE sample buffer, and the mixture heated at 90°C for 2 minutes prior to loading onto an SDS-polyacrylamide gel. The remaining cell suspension was lysed by sonication, microcentrifuged to separate the soluble and insoluble fractions and then the soluble fraction was carefully removed and filtered through a 0.2 µm

microcentrifuge spin filter (Sartorius) to ensure that there were no residual insoluble particulates present. The insoluble pellet was then resuspended in 250 µl of water, mixed in order to wash away residual soluble fraction, and harvested by microcentrifugation. The supernatent wash was removed and discarded, and the pellet resuspended in its original volume (as calculated above) of water. This process provided a suspension of insoluble protein and a solution of soluble protein (each of equivalent volume and from an equivalent density of cells). A 15 µl aliquot of each was taken, heated at 90°C for 2 minutes in 5 µl of SDS-PAGE sample buffer then analysed by SDS-PAGE.

This process allowed an assesment of protein expression with time post induction, and an indication of protein solubility at each stage following induction. Thus the optimum point for cell harvest could be determined. Where expression was poor Western blots were used to detect the desired protein (see 'Western blots and dot blots', on page 153 within the 'Protein preparation and analysis' section).

Large scale cultures

E.coli, from a glycerol stock, carrying the appropriate plasmid were spread onto an agar plate supplemented with the antibiotics required to maintain selection of the plasmid. The plate was incubated at 37°C overnight allowing a lawn of confluent cells to develop. Cells were scraped from one half of the plate (using a sterile disposable cell scraper from Costar) and used to innoculate a starter culture of 25 ml LB, also containing the necessary antibiotics. The cells remaining on the other half of the plate were used to innoculate an additional 25 ml broth, which would be used as the starter for another, concurrently prepared, main culture (this text describes the steps used in preparation of just one of these cultures). The 25 ml starter culture was incubated with shaking at 37°C[8] for 3 hours whereupon the cells were harvested by centrifugation and resuspended in 5 ml of fresh LB. The resuspended cells were used to innoculate a prewarmed (to 37°C[8]) 1 litre main culture of LB, supplemented with appropriate antibiotics.

[8] Unless otherwise stipulated

Main cultures were incubated with shaking at 37°C[8], and induced by addition of IPTG, to a final concentration of 1 mM, on reaching an $OD_{600} \approx 0.5$ - 0.7. Each culture was allowed to continue growing for the optimum period post induction, as determined by the small-scale trials (described in the previous section).

The cells were harvested by centrifugation, and resuspended in buffer **A**[9] (table 9) at a density of approximately 12.5 ml buffer per litre of cells cultured. The cell suspension was stored frozen at -70°C pending protein purification. Where IMAC was to be used as a purification procedure the EDTA and DTT were omited from the buffer. When required larger culture volumes were grown and induced accordingly.

Table 9 Buffer A.

50 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT, 1mM PMSF

Fermentation

Fermentation was often used as an alternative to shake flask cultures when large-scale (8 L) preparations were required. It offered the advantage of increased biomass, resulting largely from more efficient aeration and the use of a richer media. A disadvantage was the increased possibility of plasmid loss without continual selective pressure in favour of the plasmid maintenance.

An LH series 210 fermenter (from Inceltech) was used. This featured a direct drive motor, 10 L maximum capacity, temperature control, pH control, and provision for use of Oxygen (which was not utilised). The media used was 'Terrific Broth' (table 10) and a pH of 7.2 was maintained by the buffering effect of Potassium diHydrogen orthoPhosphate within the media.

Table 10 Terrific Broth.

Bactotryptone 20 gL^{-1}, Yeast extract 10 gL^{-1}, Sodium Chloride 5 gL^{-1}, Potassium diHydrogen orthoPhosphate 6.8 gL^{-1}, Glucose 1 gL^{-1}, and Glycerol 2 mlL^{-1} - adjust pH to 7.2 using 40% Sodium Hydroxide.

[9]Except where specific alternative buffers are stipulated

A typical run used 8 L of terrific broth. The ingredients were dissolved in 1 L, adjusted to pH 7.2, transferred to the fermentation vessel, then made up to the final volume required with water. The fermenter vessel was assembled according to the manufacturers instructions with the media *in situ*, and autoclaved at 121°C for 40 minutes at 115 psi. Autoclaving was usually carried out the night before a run, and the fermentation vessel remained in the autoclave until morning. In this way the vessel cooled slowly, and had generally reached about 37°C by the time it was required. Media and glassware for preparation of the starter culture were autoclaved at the same time.

In order to produce sufficient starter culture for the fermenter 75 µl of *E.coli* glycerol stock containing the desired plasmid was spread onto each of four agar plates supplemented with appropriate antibiotics to maintain plasmid selection. The plates were incubated at 37°C overnight, producing a confluent lawn of cells. The cells were scraped from each plate using a sterile disposable scraper and used to inoculate 100 ml of Terrific Broth also supplemented with the appropriate antibiotics. This starter culture was incubated with shaking at 37°C for three hours. The cells were then harvested by centrifugation, resuspended and washed in 100 ml LB, re-harvested and finally resuspended in 100 ml of fresh Terrific Broth containing the antibiotics. This cell suspension was then used to inoculate the main fermentation vessel, also supplemented with the appropriate antibiotics. The purpose of this lengthy inoculum preparation procedure was to obtain an actively growing culture, which had been subject to stringent selection pressure aimed at maintaining the desired plasmid, with a view to ensuring that high levels of the plasmid were present in the main culture.

The culture was grown at 37°C, with a stir speed of 800 rpm, and an air-flow rate of greater than 3000 $mlmin^{-1}$ (usually 5000 $mlmin^{-1}$). Airflow into the vessel was passed through Whatman Hepa-Cap 0.2 µm filters (previously autoclaved and oven dried) to ensure sterility of the air intake. The Hepa-Caps were chosen because they maintained a high air-flow rate, without blockage, throughout a run. The cell density (OD_{600}) was monitored throughout the run via a sample collection port, and induction was achieved by addition of IPTG to a final concentration of 2 mM (via an inlet port) once mid-log phase was reached at an OD_{600} of 0.5. The culture was then grown for a further 2 hours, and typically reached a final OD_{600} approaching 2.0. Total run times

were usually less than 6 hours. At the end of a run the heating element was unplugged and the bulk of the culture was expelled via the sample collection port, by exploiting the positive air pressure within the vessel, and collected for harvest. When no further culture could be removed in this way the vessel was disassembled and the remaining culture poured out manually and harvested. The fermenter and vessel were then cleaned and disinfected as per the manufacturers recomendations. The cell paste was weighed, typically an 8 L fermentation yielded 40 g of cell paste, and resuspended in 200 ml buffer A prior to storage at -70°C pending purification.

Lysis of bacterial cells

In order to release over expressed protein from bacterial cells prior to purification the cells must be disrupted. The French Press and Sonication techniques were routinely used, and the instruments were operated in accordance with manufacturers recomendations. Freeze thawing of cells in the presence of lysozyme was often used to augment mechanical lysis techniques.

Using a French Press

The French press comprises a stainless steel cylinder closed at one end, except for an adjustable narrow bore aperture, and fitted with a stainless steel piston from the other end. The whole assembly sits in a large motorised press which allows the operator to lower or raise the piston. The cell slurry to be disrupted is poured (or drawn via an inlet port) into the cylinder, and then forced out through the narrow aperture as pressure is applied via the piston. A combination of increased pressure, followed by rapid decompression, and shearing forces results in cell lysis. It was often necessary to perform repeated passes to ensure complete lysis. The cylinder assembly was chilled at 4°C prior to use, and sample slurries were kept on ice throughout the process. The main advantage of the French Press method was that no heat was generated during the process, reducing the likelihood of unwanted proteolysis or denaturation of protein.

Using a Probe Sonicator

Ultrasonication involves the exposure of cell slurry to high frequency vibrations (20 kHz or more) generated by a sonication probe immersed in the slurry. The high frequency vibrations cause areas of rapidly interchanging compression and rarefaction (gaseous cavitation), which in turn generate shock waves able to disrupt bacterial cell walls. Typically the cells for disruption were resuspended in lysis buffer at 20% w/v or less (to reduce viscosity and ensure adequate mixing). The buffer used varied according to the protein to be isolated, but usually incorporated 1 mM EDTA and 0.1 mM PMSF to reduce proteolysis. Disruption is dependent on exposure time, however, a great deal of heat can be generated during the process which will cause proteins to denature. For this reason the sonicator probe was pre-chilled by storing on ice for 20 minutes prior to use, and cell slurries were kept on ice throughout the procedure. Also sonication was controlled using a programmable timer and carried out in short bursts (30 secs each), interspersed by periods of cooling (1 or 2 minutes). Sonication was especially used for disrupting small volumes, such as 1 ml slurries in (an eppendorf) for small scale protein expression analysis.

By freeze thawing in the presence of lysozyme

This method of cell disruption was often used in conjunction with sonication in order to reduce the sonication time required for complete disruption. Lyophilised hen egg white lysozyme was dissolved in the buffers used for cell resuspension following harvest. The cell pellets were stored frozen until required, then thawed on ice. During the thawing process the lysozyme begins to break down the peptidoglycan in the bacteria cell walls. A 50 ml cell paste typically took 30 minutes to thaw, and in this time the majority of cells would have lysed. The thawed paste was usually very viscous due to the release of DNA from the cells, this was reduced by the addition of DNAse and RNAse to the resuspension buffer following harvest, or by a subsequent sonication step. Cells that expressed lysozyme (eg cells containing the pLys S plasmid) obviated the need for inclusion of hen egg white lysozyme in the harvest buffer.

Protein preparation and analysis

SDS-PAGE Analysis of Proteins

Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE) of proteins was carried out using methods based on those described by J. V. Maizel and co-workers, and Laemmli (Shapiro, Vinuela *et al.* 1967; Maizel and Summers 1968; Shapiro and Maizel 1969; Laemmli 1970). The apparatus used in all cases was the BioRad Miniprotean system, and this was assembled and used as per manufacturer's instruction. Resolving gels were typically either 12% or 15% with respect to acrylamide content, as required. The constituents of gel mixes (stacking and resolving), loading (sample) buffer, and running buffer are given in the table below. Gels were electrophoresed by application of 30 mA current per gel for a period of approximately one hour (depending on the specific requirements of the experiment), using a BioRad 'Power Pac 300' (*sic*) power supply. Samples containing high salt, urea or guanidine were first acetone precipitated, then resuspended in 1 x loading buffer. This technique avoided poor resolution and other problems commonly associated with high salt samples, and also obviated the need for gels containing chaotropic reagents.

Size Standards for PAGE

Standard size markers were used with each gel to provide an estimate of protein size, and also to indicate successful electrophoresis. The markers used in each case are stipulated in the text. Prestained markers were often used since they allowed a real time indication of the extent of electrophoresis during a run, and were also useful indicators for Western blots.

Acetone precipitation

An aliquot of up to 250 µl of the protein solution to be precipitated was taken in a 2 ml eppendorf, which was then filled to the maximum fill line with acetone. The contents of the tube were mixed by inversion or vortexing, and then subjected to microcentrifugation at maximum rpm for at least 10 minutes, usually 30 minutes. The supernatent was discarded, and the precipitate allowed to air dry prior to resuspension as desired.

Western blots and Dot blots

Blotting was first described as a technique for the transfer of DNA from agarose gels to nitrocellulose membrane (Southern 1975) and was later applied to the transfer of proteins (Renart, Reiser *et al.* 1979; Bowen, Steinberg *et al.* 1980). Electrophoretic transfer as a means of eluting protein from a gel onto a membrane was first described by Towbin *et al.* (Towbin, Staehelin *et al.* 1979), and is known as 'Western blotting'.

Proteins are first separated according to their molecular weight via conventional PAGE (as described above), then transferred horizontally onto a nitrocellulose membrane by electrophoresis. The membrane is then 'blocked' by incubation with a solution of BSA or skimmed milk powder to prevent non-specific binding of antibody to the nitrocellulose. The membrane is then exposed to a primary antibody specific for the protein of interest. A secondary antibody is then applied which specifically recognises the primary antibody, and is labelled (either chemically, photochemically, or radioactively) to allow its detection. Finally binding of the secondary antibody is detected (method dependant on the label used) and indicates the presence (or absence) of the protein of interest.

The Western blots described here were carried out using the BioRad 'Mini Trans-Blot' electrophoretic transfer cell, which is compatible with the 'Mini-PROTEAN' system used here for PAGE, and also supplied by BioRad. Details of the assembly and operation of the blotting apparatus are available from the Mini Trans-Blot Instruction Manual, supplied by BioRad. Details of the assembly of blot components within the cassette are given below. Two transfers could be carried out simultaneous, running them at 400 mA for 90 minutes.

Table 11 Western blot transfer buffer.

25 mM Tris pH 8.3, 192 mM Glycine, 20% v/v methanol
It should be noted that this buffer was used with nitrocellulose membranes. Other media may require different buffer.

Dot blots were developed in the same way as Westerns, but differed in that protein was placed directly onto the membrane rather than transferred from a preceding PAGE step. Dot blots provided a quick visual assay for the presence of a specific protein in a solution, but did not give any indication of the relative size of the protein in question.

Assembly of the blotting components within the BioRad cassette.

A PAGE gel was run, as previously described, using prestained markers. The gel was carefully removed from between the glass plates and placed into a dish of Western blot transfer buffer (table 11). For each gel to be transferred a square of nitrocellulose membrane was cut to size (approximately 7.5 cm x 10 cm) and soaked in transfer buffer, as were two 7.5 cm x 10 cm squares of Whatman no.1 filter paper, and two fibre pads (supplied with the blotter).

The blotting cassette was placed open in a shallow dish filled with tranfer buffer, with the grey panel (to be orientated towards the cathode during the run) in the base of the dish. Gloves were worn throughout component assembly and disassembly, to avoid contaminating the blot, or contact with polyacrylamide. A pre-soaked fibre pad was placed on the grey panel, followed by a saturated square of filter paper, and then the gel, with its surface flooded with transfer buffer. The nitrocellulose membrane was next laid on top of the gel and a glass pipette was rolled over the membrane to exclude any trapped air bubbles from between the gel and the membrane. The surface of the membrane was then flooded with buffer, and a square of filter paper placed over it, followed by a second fibre pad. When assembling the blot components great care was taken to ensure that no air bubbles were incorporated within the layers, and that they were centred within the cassette. This ensured a complete and undistorted transfer. The closed and secured cassette was then placed within the electrode assembly, orientated such that the grey panel (i.e. the gel side) faced cathode electrode.

Blot development

Following transfer (or dot blotting) the nitrocellulose membrane was incubated with shaking in blocking solution (table 12), at 4°C for at least 30 minutes, then washed twice - first with PBS + 0.1% Tween 20, then with PBS alone.

Table 12 Western blot blocking solution

Phosphate buffered saline (PBS), 5% w/v dried milk powder *, 0.1% TWEEN 20 PBS tablets are available from Sigma, use one tablet per 200 ml of solution. (* marvel or five pints etc.)

The membrane was then incubated in PBS containing 5% milk powder and primary antibody, for at least one hour - or even over night. The antibody was used at the final concentration recommended by suppliers. Incubation times required vary with the efficacy of the antibody, and with levels of the target protein present. Here the membranes were exposed to primary antibody overnight, since protein expression levels were usually poor. The primary antibody solution can be re-used several times, and was stored at -20°C between blots.

After exposure to the primary antibody, the membrane was washed again in two changes of PBS / TWEEN, then two changes of PBS. Next the membrane was exposed to the secondary antibody in PBS for one hour (the secondary antibody solution was not re-used). The PBS / TWEEN and PBS wash steps were repeated, and the membrane finally washed in two changes of distilled water.

The secondary antibodies used were conjugated to alkaline phosphatase, consequently blots were developed by incubation in a 10% solution of 1,4-dichloronapthol, until bands appeared and were fully developed. The reaction was stopped by removing the blot from the developer and immersing in distilled water, then changing the distilled water several times to ensure thorough washing. The membranes are extremely fragile - especially when dried - and also fade or discolour with age. They can be dried between transparent sheets (using the BioRad gelair system) in the same way as protein gels, and this helps prevent deterioration.

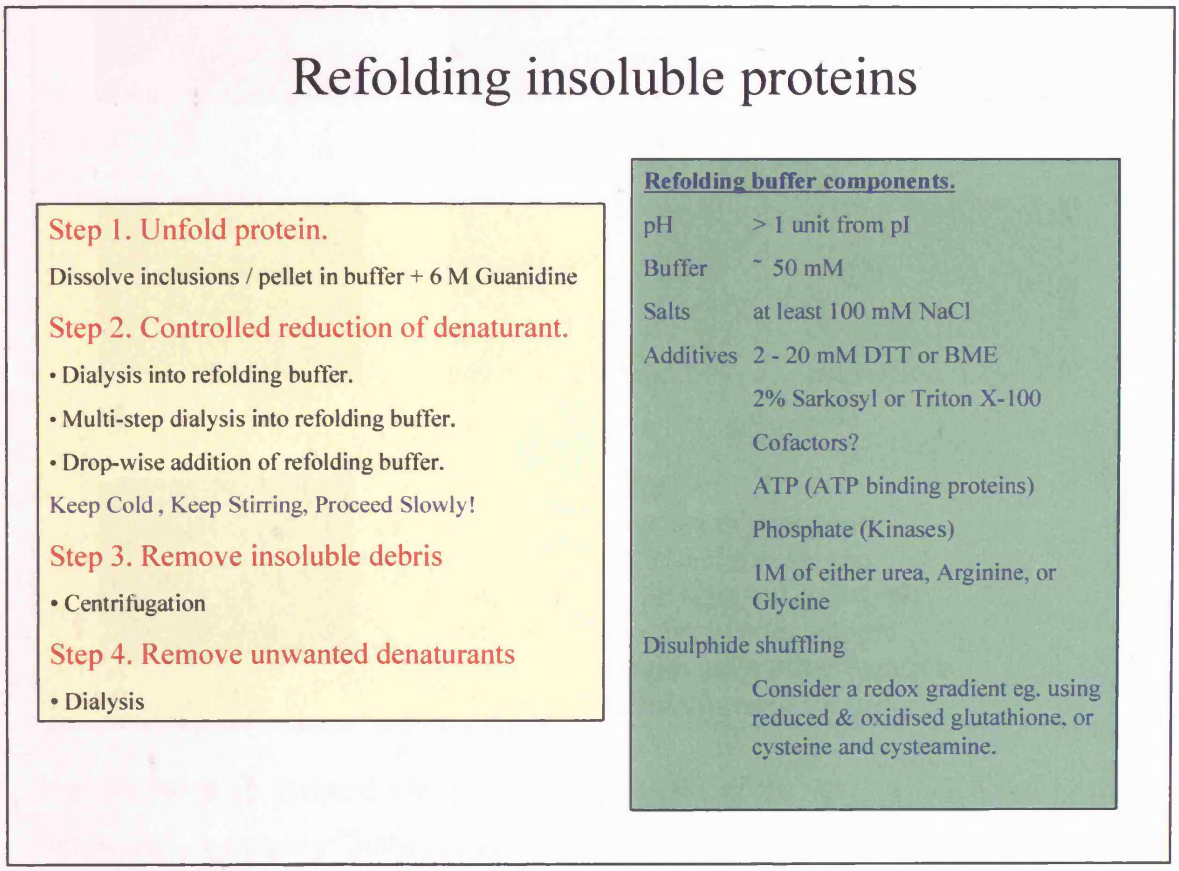
Refolding insoluble proteins.

Where protein is expressed in the insoluble fraction of the cell lysate (as inclusion bodies) it can be isolated by centrifugation (retaining the pellet) and in many cases refolded. Frequently insolubility is caused by mis-folding of the protein during expression, resulting in deposition as inclusion bodies. It is usually better to optimise soluble expression rather than resort to refolding, however, occasionally this is either unsuccessful or undesirable (eg. when deposition of a protein in inclusion bodies is intentional). In these cases refolding will be necessary. Following complete cell lysis the insoluble fraction is isolated as a pellet by centrifugation, and washed with a low chaotropic buffer such as 0.5 M guanidine or Urea. The pellet is then dissolved in high concentration chaotropic buffer (eg. 6 M guanidine or 8 M Urea), gentle sonication and stirring may be required to ensure total dispersion of the pellet. Any remaining insoluble particles are then removed by ultracentrifugation, and the clarified supernatant retained. This contains the unfolded protein. Refolding is achieved by gradually reducing the levels of chaotrope until they are negligeable. The refolded protein should remain in solution, but inevitably some components of the solution will precipitate from solution and these are removed by ultracentrifugation. A concentration step will then reduce the volume of the refolded protein solution and concentrate the target protein.

Gradual removal of the chaotrope can be achieved in several ways, which is chosen depends on the ease with which a protein refolds. The quickest and simplest method involves conventional dialysis into a non-chaotropic buffer; this is the method of choice. Where straight forward dialysis is unsuccessful step dialysis against a series of progressively lower chaotropic refolding buffers may be employed. Finally, for proteins that do not refold well by dialysis, a slow drop-wise dilution of the solution with a non-chaotropic refolding buffer can be used. For all methods, the components of the refolding buffer should be optimised for solubility of the target protein by careful choice of characteristics such as pH, salt concentration, and presence of additives (such as glycerol, DTT or cofactors). All steps in the refolding process should be carried out at 4°C.

The retroviral proteases are relatively small proteins, and refold very easily - usually by means of a single dialysis step. Refolding was thus used during several stages of protein purification for each of the proteases described in this thesis. Specific methods and buffers used for each protease are described in the methods sections of the relevant chapters. Figure 1 summarises the basic methodology for protein refolding in general, and the characteristics of refolding buffers that must be considered.

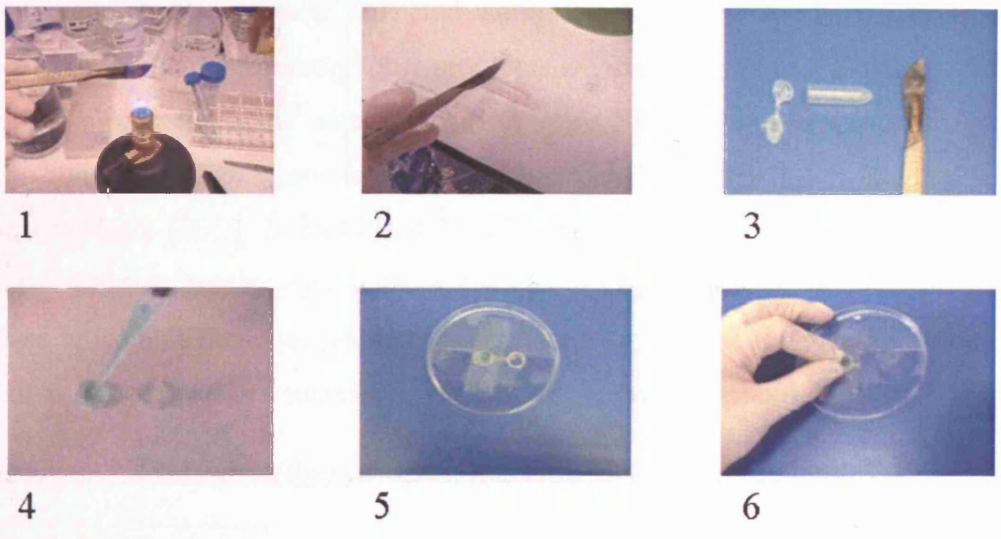
Figure 1 The refolding of insoluble proteins.



Microdialysis.

Microdialysis was used where small (µl) volumes were to be dialysed – for example during testing of inhibitors in Chapter 3, or the refolding steps described in Chapter 6 for preparation of SIV_{AGM} PR. Microdialysis buttons were generated from the tops of eppendorf tubes, as shown in figure 2.

Figure 2 Production of microdialysis buttons.



1. Heat scalpel
2. Cut below rim of eppendorf
3. Remove lid and rim for use
4. Fill lid with sample for dialysis
5. Place membrane over sample
6. Hold membrane in place with rim
7. Dialysis button ready for use

The sample to be dialysed was pipetted into the lid of the tube, which acted as the button well. A square of dialysis membrane was placed over the sample and then held in place by fitting the rim of the tube over the lid. The completed button was then placed into the dialysis buffer, membrane downwards, for dialysis. Once dialysis was complete, the sample was removed by slitting the membrane with a scalpel, pushing a pipette tip through the slit, and drawing the sample out. This system was inexpensive, used readily available materials, and was highly effective.

Chromatography

All chromatographic procedures were carried out using a Fine Performance Liquid Chromatography (FPLC) system from Pharmacia (subsequently Amersham Pharmacia Biotech, now GE Healthcare), and all columns and packing materials were supplied by Pharmacia unless otherwise stipulated. Chromatography was carried out at room temperature, unless stated otherwise, and all buffers were drawn through a 0.2 μm filter by vacuum prior to use. Eluted protein was detected by UV absorbance at 280 nm, and fractions were collected automatically using a Pharmacia fraction collector. Individual chromatographic protocols are described in the relevant text. The modes of chromatographic separation used were Size Exclusion Chromatography (SEC), Hydrophobic Interaction Chromatography (HIC), Ion Exchange Chromatography (IEC), Immobilised Metal Affinity Chromatography (IMAC), and Affinity Chromatography - particularly the use of Pepstatin Agarose. The principles involved in each of these techniques are explained well in the texts below (table 13), and are also explained in a number of general biochemical textbooks.

Table 13 Texts describing chromatographic analysis of proteins.

• Protein Purification, Principles and Practice (Robert K Scopes)
• Protein Purification, A Practical Approach (T.E. Creighton)

Pepstatin-agarose affinity chromatography

Pepstatin-agarose resin (obtained from Sigma) binds aspartic proteases generally, with a binding capacity of approximately 1 mg porcine kidney renin per ml of resin. Since pepstatin is a potent inhibitor of both HIV and SIV proteases this resin allowed affinity purification of the retroviral proteases studied here. The principle of using a peptide substrate analogue immobilised on agarose resin for the purification of HIV proteases has previously been described by Heimbach *et al.* 1989 (Heimbach, Garsky, *et al.* 1989) and protocols for the use of pepstatin-agarose have been described by Rittenhouse *et al.* 1990 (Rittenhouse, Turon, *et al.* 1990) and Wondrak *et al.* 1991 (Wondrak, Louis, *et al.* 1991). In general pepstatin inhibition of HIV-1 PR is greater at high ionic strength, and optimal below pH 5.5 becoming weaker above pH 6 (Billich *et al.* 1990). Elution can therefore be achieved by means of a reduction in ionic strength and an increase in the pH.

The purification procedure adopted was essentially as described by Rittenhouse *et al.* 1990. Pepstatin-agarose was packed into a 5 ml column (Pharmacia), and washed thoroughly with equilibration buffer (table 14). Samples were prepared in equilibration buffer and applied to the column, the flow through was collected for later analysis. The column was then washed with five volumes of equilibration buffer.

Table 14 Pepstatin-Agarose equilibration buffer.

50 mM HEPES pH 7.5, 1 mM EDTA, 1 mM DTT, 0.1 mM PMSF

The elution buffer (table 15) was applied to the column, and 10 ml fractions were collected into tubes containing 0.5 ml 2M HEPES, immediately neutralising the samples. Fractions collected in this way were stored at –20°C pending further use.

Table 15 Pepstatin-Agarose elution buffer.

250 mM Sodium ε-aminocaproate pH 10.5, 5% glycerol, 5% ethylene glycol, 1 mM EDTA, 1 mM DTT, 0.1 mM PMSF

The column was cleaned immediately after use with five volumes of 6 M guanidine-HCl followed by five volumes of equilibration buffer containing 20% ethanol. All purification steps were carried out using a Pharmacia FPLC, at a flow rate of 0.5 ml/min.

Concentration of proteins by ultrafiltration

Ultrafiltration (UF) uses semi-permeable membranes to separate solvents from solutes of various sizes. UF membranes have a thin dense skin with pores in it (whose size are precisely controlled during manufacture), supported by an open-celled spongy substructure. Any molecular species able to pass through the pores of the skin will be able to freely pass through the substructure. Thus retention of a molecule is a function of its size and shape, compared with the pore size of the membrane. Membranes are available with various pore sizes, and thus varying degrees of solute retention. They are assigned a molecular weight cut-off (MWCO) in Daltons to indicate the molecular weight of solute which will be 90% retained by the membrane. Any molecule smaller than the MWCO for a given membrane may pass through that membrane. The membranes are inert, non-cytotoxic, do not bind or denature proteins, and withstand repeated use (provided the outer skin is not damaged). UF was predominantly used as a means of concentrating protein samples, but could also be used to desalt or buffer exchange samples.

Pressure cells and centrifugal concentrators were used as described below. In both cases membranes were chosen according to the size of protein to be concentrated, and were prepared and used in accordance with the manufacturers instructions - with particular attention to pressures or RCF applied.

Amicon pressure cell concentrators

Large volume protein samples (ie. ≥ 100 ml) were reduced in volume using UF in a stirred pressure cell supplied by Amicon. The protein solution to be concentrated was held within the body of the pressure cell, and gently stirred throughout the concentration process by a magnetic stirrer suspended from the cap assembly of the cell. The membrane was supported by a holder which sat in the base of the cell, and leaks were prevented by O-rings in the cap and base assemblies. The entire cell was positioned inside a retaining stand, which prevented the cap from blowing off once gas pressure was applied. Nitrogen gas was then applied through a nozzle in the cap assembly, and the pressure exerted by the gas forced solvent through the membrane at

the base. Solute molecules larger than the MWCO for the membrane used were retained within the cell body. The stirrer prevented concentration polarisation (which would have reduced the flow rate). Amicon pressure cells of various volumes were used, depending on initial sample size, and concentration was carried out at 4°C. Care was taken to ensure that the cells did not run dry, and that precipitation of the sample protein was avoided. Cells were assembled and operated as recomended by the manufacturer.

Centricon and Centriprep spin concentrators

These UF protein concentration devices were also supplied by Amicon, and were used to concentrate small sample volumes (10 ml - 100 µl). Protein solution was held in a reservoir above the membrane, which was itself supported above a flow through collection tube. The entire assembly was spun in a fixed angle centrifuge rotor and solvent was pushed through the membrane by centrifugal force. The maximum RCF that could be applied during a spin was determined by the membrane type, and was specified in literature accompanying the concentrators. During the concentration process the centrifuge was periodically stopped so that the speed of progress could be monitored. Sample could be topped up if necessary, and the flow through tube emptied if required. The sample in the reservoir was regularly mixed using a pipette to avoid concentration polarisation. Periodic inspection of the sample also reduced the risk of sample precipitation or of the concentrator running dry. Following concentration the flow through collection tube was removed, and a retentate collection tube was fixed over the opening of the sample reservoir. The concentrator was then inverted and spun at a low RCF to collect the concentrated sample solution.

Protease assays

Monitoring protease activity.

Protease activity was monitored where possible using an assay developed commercially by Amersham International plc (subsequently Amersham Pharmacia Biotech, now GE Healthcare). This is a system based on the scintillation proximity assay (SPA) principle (Udenfriend, Gerber *et al.* 1987; Cook, Jessop *et al.* 1991) whereby beads containing a scintillant are attached to radioactive ^{125}I via a peptide, which acts as the proteolytic substrate. The proximity of ^{125}I to the scintillant causes the beads to emit light which can be detected by a scintillation counter. If proteolytic activity severs the peptide, causing the ^{125}I and the bead to separate, scintillation will cease. Protease activity is therefore detected as a drop in scintillations counted. This assay was used as a rapid means of following the protease activity during the purification procedures. Assays were routinely carried out at 37°C for 1 hour. Typical reaction conditions consisted of 100 µl SPA reagent, 10 µl enzyme, and 90 µl SPA reaction buffer (table 16).

Table 16 SPA reaction buffer

100 mM 2-(N-Morpholino)ethanesulphonic acid (MES) pH 6.5, 1 mM DTT, 1 mM EDTA Plus 200 mM NaCl for HIV-2 PR assays (not required for HIV-1 PR assays).

Kinetic analysis of proteases

The SPA assay was unsuitable for kinetic studies, therefore K_M determinations were carried out using chromogenic substrates (supplied by Bachem) under the conditions described in Pennington *et al.* (1990) (Pennington, Dick *et al.* 1990). This type of assay exploits the difference in molar absorptivity between the substrate and the products, where the change in absorption is proportional to the extent of cleavage. It is preferable to use a system where this difference is greatest in the visible range, since measurements in the UV range will be subject to interference from other components of the sample extract (especially if crude lysates are analysed). Details of substrates, precise conditions, and instrument settings used are given in the relevent text.

Data was recorded using a Shimadzu Graphicord UV-240 dual beam spectrophotometer (able to monitor absorbance over a range from 190 - 1100 nm) with PR-1 printer attached. Samples were analysed using a pair of matched 1 ml quartz cuvettes plus lids (supplied by Hellma). Reaction buffers varied dependent on the protease under study, and are described in the relevant text. All buffers were prewarmed to 37°C, while enzyme and substrate were stored on ice prior to use. The spectrophotometer's chamber was maintained at 37°C by water pumped from a 37°C bath through heating ducts within the chamber. Final reaction volumes were 1 ml and measured against a blank containing buffer alone. Addition of substrate and enzyme to the reaction mix was compenstated for by prior removal of an equivalent volume of reaction buffer, so that the 1 ml final volume was always maintained. Substrates were supplied as a lyophilised powder, and were dissolved in water to give a concentrated stock solution which was aliquoted and stored at -20°C. An aliquot of this stock was thawed on ice immediately prior to use, and diluted with water to provide sufficient substrate of the required concentration for that day's experiments.

Table 17 Procedure for protease assays using chromogenic substrate.

1.	The spectrophotometer housing was allowed to warm to temperature (37°C), and the appropriate wavelength (300 nm), chart speed (20 nm/cm), scan speed (slow), mode (time), and deflection settings (a maximum deflection of 0.05) were set.
2.	Sample and reference cuvettes were filled with buffer alone, and the spectrophotometer was adjusted to zero.
3.	Substrate stock (volume calculated to achieve the desired final concentration) was added to the sample cuvette and mixed by inversion. A base line measurement was then made.
4.	The enzyme (usually 10 - 20 µl) was added to the sample cuvette, mixed, and the initial velocity of any proteolytic activity measured.

Each reaction was measured for a sufficient length of time to allow a reasonable linear trace to be obtained. If the slope of the trace was too great (ie greater than about 60°) adjustments were made to rectify this. Typically either the spectrophotometer's deflection was adjusted, a faster chart speed was selected, or the enzyme stock was diluted. Readings were taken in triplicate for a given substrate concentration, their initial velocities were calculated and a mean value determined. Results were obtained over a range of substrate concentrations (stipulated in the relevant text) and values for K_M, K_{cat} and V_{max} were derived using a computer fit of the data calculated by the Enzfitter program (from Biosoft).

Small scale affinity purification experiments.

General principle

These experiments exploited the ability to isolate a protein on a bead by means of a fusion tag or by affinity purification. They were carried out on a small scale (~ 100 µl volume) in eppendorf tubes. In principle the procedure was the same for all pull-down experiments. First the protein solution (usually 100 µl) was mixed and incubated with the chosen beads in an appropriate binding buffer. Next the beads were isolated, either by microcentrigugation or microfiltration, and the supernatent / filtrate was discarded. The beads were then washed with several volumes of fresh binding buffer to remove any unbound, or non-specifically bound protein. Finally the beads were collected in microcentrifuge spin filters (available from Sartorius) and bound protein eluted from them into an eppendorf tube either by using an appropriate elution buffer, or by direct application of SDS-PAGE sample loading buffer. The eluted protein was then analysed by SDS-PAGE. The controls required for the gel included the input protein solution, the beads alone, and the beads plus a similar protein solution which does not include the target protein. If the target protein is present in the initial solution and has interacted with the beads it should be present in the eluted sample, whilst other proteins present should not bind to the beads.

His-tag pull-downs

A target protein having an N-terminal 6 x histidine tag was used, and pull-down was achieved using Talon resin (from Clonetech).

Table 18 His-tag binding / wash and elution buffers.

Tris pH 8, 100 mM NaCl, 1 mM immidazol Supplemented with 100 mM immidazol for elution.

MBP pull-downs

The target protein was fused to Maltose Binding Protein (MBP), by cloning into the pMalcRI vector from NEB. The resin used was Amylose (available from NEB).

Table 19 MBP binding / wash and elution buffer.

20 mM Tris pH 7.4, 200 mM NaCl, 1mM EDTA, 1mM DTT Supplemented with 10 mM Maltose for elutions.

Pepstatin-agarose pull-downs

Pepstatin immobilised on the resin binds aspartic proteases, allowing them to be pulled-down. Binding and elution are pH and salt dependent (see 'Protein preparation and analysis' - 'Pepstatin-agarose affinity chromatography' page 160).

Table 20 Pepstatin-Agarose equilibration, binding and wash buffer.

50 mM HEPES pH 7.5, 1 mM EDTA, 1 mM DTT, 0.1 mM PMSF

Table 21 Pepstatin-Agarose elution buffer.

250 mM Sodium ε-aminocaproate pH 10.5, 5% glycerol, 5% ethylene glycol, 1 mM EDTA, 1 mM DTT, 0.1 mM PMSF

Chemical characterisation

Amino terminal analysis (N-terminal analysis)

N-terminal analysis by means of automated Edman degradation reactions was carried out by the Wellcome Research labs, Beckenham.

Mass spectrometry (MS)

Electrospray Mass spectroscopy of HIV proteases was carried out by the Wellcome Research labs, Beckenham, MS analyses of inhibitor compounds were carried out by Dr. J. Hill of the Chemistry Department, UCL.

Circular Dichroism (CD)

CD analyses were carried out by Dr. Ronan O'Brian at the Biochemistry and Molecular Biology Dept. UCL, Gower St. London. For details of CD methodology, and of the interpretation of spectra see 'Protein secondary structure and Circular Dichroism: a practical guide' (Johnson 1990).

(For further reading see the following references:- (Rosenkranz 1974; Rosenkranz 1974; Johnson 1988; Johnson 1990; Bloemendal and Johnson 1995; Kuwajima 1995; Woody 1995; Greenfield 1996; Woody, Sugeta *et al.* 1996).

Nuclear Magnetic Resonance (NMR)

NMR spectroscopic analysis of the inhibitor compounds was carried out by Dr. Abil Aliev at the Chemistry department of UCL.

Crystallographic methods

Preparation of proteins for crystallography

When preparing a protein for crystallography the aim was to obtain a stable solution of pure protein at high concentration (10-20 $mgml^{-1}$). Purification of the protein to be crystallised (preferably to homogeneity) was followed by dialysis into the desired buffer for crystallisation. Dialysis included at least two changes of buffer, approximately twelve hours apart. A buffer was chosen which maximised the protein's stability in solution. If a prefered buffer had not yet been established then a 50 mM Tris-HCl buffer at pH 7.5 was used as an initial choice. Low levels (0.02%) of sodium azide were included to prevent fungal or bacterial growth during storage or crystallisation trials. In general the ionic strength of the solution was kept as low as possible without jeopardising protein stability (some proteins require the presence of salt if they are to remain in solution). Following dialysis the protein was concentrated using either an Amicon stirred pressure cell, or a Centricon (depending on the initial volume of solution). Where possible the solution was taken to a slightly higher concentration than intended, so that the concentrator membrane could be washed with buffer and the wash then incorporated into the sample (bringing it back to the intended concentration). Concentration was determined by measuring absorbance at 280 nm since it was a quick and easily reproducible method found to give a reasonable approximation to the actual protein concentration.

Once concentrated the protein was aliquoted into appropriate sized fractions (typically 100-200 μl) and stored at -70°C. Aliquoting fractions reduced the number of freeze thaw cycles the protein was exposed to, and a 100 μl aliquot was just enough to allow both Hampton sparse matrix screens to be set up.

Methods of crystal growth

The Microbatch method

This technique involves the use of microtitre plates filled with mineral oil, and drops are set up in the wells beneath the oil layer (Lloyd, Brick *et al.* 1991; Pav, Lubbe *et al.* 1994; Zhao, Smith *et al.* 1996; Chayen 1997; Chayen 1998; Dong, Boggon *et al.* 1999). The technique uses only small volumes of buffer solutions (typical drop sizes are 1 or 2 µl), can easily be automated (Chayen, Shaw *et al.* 1994), and the final concentration of drop components can be precisely determined. Once set up the screens are quite robust and trays can be transported without great risk, however, manipulation of crystals may be slightly impeded by the oil layer. The technique was predominantly used for initial screening (Sparse Matrix screening) as described in the next section.

The 'Hanging Drop' or 'Vapour Diffusion' method

This method involves the suspension of the crystallisation drop above a reservoir full of the crystallisation buffer. It exploits the dynamic equilibrium between the drop, the well buffer and the vapour in the intervening space. The drop was formed in the centre of a siliconised glass coverslip (usually 50/50 protein solution to crystallisation buffer), and was typically 2 - 4 µl in volume. A limbro plate with large wells was used to provide the reservoir which was filled with 1 ml of the crystallisation buffer. The rim of the reservoir was smeared with silicon grease, and the coverslip was inverted and carefully lowered over the reservoir so that the drop was suspended above the well. The coverslip was gently pressed into the grease to ensure a good seal. The entire plate was then incubated at the required temperature (usually 14°C for retroviral proteases), and the drop inspected by microscope at daily intervals for signs of crystal growth. Hanging drops allow a gradual drift in buffer / protein concentration within the drop as the system approaches equilibrium. Ideally a drop should pass through conditions that allow both nucleation and crystal growth as it equilibrates. Vapour diffusion is considered in greater detail in the texts recomended for further reading in the appendix (Appendix: Further Reading; Table 9, page 352).

Determination of crystallization conditions

'Wild Screens' or 'Sparse Matrix' screens

The initial conditions for crystallization were determined by means of a 'Wild Screen' or 'Sparse Matrix' approach (Jancarik and Kim 1991). This employs a total of 50 buffered, precipitant solutions selected from known or published successful crystallisation conditions. This principle has subsequently been used to create various commercially available screens. The solutions used here were initially prepared in the laboratory at UCL in accordance with Jancarik and Kim 1991, though the 'Crystal Screen I' by Hampton Research was subsequently used. The screen was carried out using droplets under oil, in microtitre plates, as described in the previous section (ie. by microbatch). Further details of the screen are available at www.hamptonresearch.com/2110.html.

Optimisation

Once an initial condition had been established, crystals were grown by the hanging drop vapour diffusion using that same condition. The first few drops were observed carefully and further drops were set up using subtle variations of the initial conditions to achieve optimum crystal growth. In particular variations in protein and precipitant concentrations were explored by titration over a series of drops. The ultimate aim was to produce good crystals of a reasonable size (preferably greater than 100 μm in diameter) which would be suitable for data collection. Establishing a crystallisation condition can be a complex, lengthy procedure, for more detailed information and methodology see 'Protein Crystallization: Techniques, Strategies, and Tips – A Laboratory Manual (Bergfors 1999).

Macroseeding

In circumstances where large individual crystals are difficult to obtain, a potential strategy is to seed fresh drops with a single crystal harvested from a previous drop. The seeded crystal acts as a nucleation site, and with luck it will grow at the expense of any other nucleation sites in the drop. The composition of the drop should set such that in the absence of a seed nucleation would not occur, but in the presence of a seed crystal growth will occur. This strategy is termed macroseeding because the nucleation site (seed crystal) is already fairly large (ie visible under the microscope and identifiable as a crystal). Other seeding techniques involve the use of seeds that cannot necessarily be seen, and tend to be used where nucleation events are otherwise rare. Further information on seeding techniques can be found in 'Protein Crystallization: Techniques, Strategies, and Tips – A Laboratory Manual' (Bergfors 1999). Seed crystals were grown under hanging drop, as described. Then a 'pre-quilibrated' buffer was prepared by pooling the well buffer from previously used hanging drop wells and 0.2 μm filtering it. It was reasoned that the constituents of this buffer would be equivalent to those of the droplet in which the seed crystals had grown. Consequently this solution could be used for the manipulation, relocation and seeding of crystals without risk of their dissolution. Hanging drops were set up using part fresh protein solution, part pre-equilibrated buffer, and were seeded with a single crystal. Crystals to be seeded were transferred direct from their original drop, which was first swelled using pre-equilibrated buffer or the drop's own well buffer, to the new drop using a Gilson P10 pipette. It was found, contrary to textbook recomendations, that the washing of crystals prior to transfer was counter productive, and caused the 'showering' it had been intended to prevent. This was due to the extreme delicacy of the crystals, any excess manipulation caused minute fragments of crystal to break away and act as nucleation points for crystalisation. This was undesirable because it deprived the main crystal of the additional protein intended to increase its size. Since the laboratory temperature fluctuated greatly, and occasionally caused unwanted evaporation of droplets during manipulation, a procedure was adopted to minimise such loss to the atmosphere. A petri dish containing a ring of Whatman no.1 filter paper soaked in distilled water was used such that a coverslip supporting a drop to be manipulated could be positioned in the centre of the ring. This provided a humid microatmosphere for the drop during manipulation, and if required the petri dish lid could be replaced allowing the drop to be left temporarily without

risk of evaporation. This was also useful when studying or manipulating a drop under the microscope, when the light source may be developing significant heat.

Three well buffer types were used with macroseeded droplets:

A. Fresh buffer as for a normal drop

B. Pre-equilibrated buffer (ie well buffer already used to grow crystals)

C. Fresh buffer with a slightly increased concentration of the various constituents intended to concentrate the drop further thereby inducing additional deposition of protein onto the seed crystal.

Mounting crystals for X-ray analysis

Crystals were mounted in thin walled glass capillaries (supplied by Hampton) suitable for X-ray data collection. The thin glass minimised both absorption of scattered X-rays and background due to the glass. These capillaries were supplied sealed at both ends, so that the ends had first to be snapped off using forceps (whilst trying to avoid the inclusion of minor glass fragments into the capillary) - this also served to shorten the capillary, since most are too long to fit on an X-ray camera unmodified. To avoid shattering the capillary during this step the outside was coated in bees wax at the point where it was to be snapped. The capillary could then be attached via fine rubber tubing to a Hamilton syringe, enabling a suitable buffer (any buffer in which the crystal remained stable - for the purposes of this work the crystallisation buffer itself was used) to be sucked part way into the tube. In this way a reservoir of buffer was created inside the capillary once the tube was sealed, providing a stable atmosphere to maintain the integrity of the mounted crystal. The syringe was then removed from the capillary. The capillary was then attached to a microscope slide, using plasticine, with the opposite end to the reservoir upmost. This enabled the crystal to be deposited into the end of the capillary. The hanging drop containing the crystal to be mounted was swelled in volume using its own well buffer. The crystal was then picked up in a small drop of buffer using a Gilson P10 pipette, and transferred directly into the capillary. This was found to be easier than alternative methods which involved using a Hamilton syringe to suck the crystal up from its drop into the capillary. A syringe was occasionally used to draw a crystal further into its tube once deposited there, since the aim was to mount the crystal roughly midway into the tube. The liquid surrounding the crystal would then be reduced or removed using a fine length of twisted lens tissue to wick it away. Excess liquid surrounding a crystal would otherwise result in a large solvent ring during data collection, and may also have allowed the crystal to slip during the collection process. With the crystal in position and a reservoir of buffer at one end (to provide a saturated atmosphere within the sealed capillary), the tube was sealed at each end using melted bees wax. The approximate position of the crystal was marked on the outside of the tube using a fine tipped indelible marker. This enabled the crystal to be found more easily when aligning it for data collection. A crystal mounted in this way can be seen in chapter 4 figure 15 (page 226).

Data collection

The basic instrumental components involved in data collection.

- The source. A source of X-ray radiation (electromagnetic waves of lengths between 0.1 and 100 Å).
- The optics (monochromator plus collimator). A means of creating a tight X-ray beam of the desired wavelength from the source.
- The goniometer and goniometer head. Apparatus to hold and manouvre the crystal in the beam.
- The beam stop. A shield to prevent the main beam passing straight through from the back of the crystal to the detector.
- The detector. A means of detecting the diffracted X-rays scattered by atoms within the crystal.

Crystal alignment

During data collection X-rays are passed through the crystal, and at the same time the crystal may be rotated in the beam. Maintaining the crystal in the centre of the beam is achieved by first attaching the capillary in which the crystal is mounted to a goniometer head. This accessory to the camera allows both rotation and translation of the crystal with respect to the beam by means of a goniometer (a system of stepping motors). The capillary is held in place on the goniometer head using plasticene, and the gonoimeter head is attached to a stand, called an optical analyser, that emulates the rotation of the camera whilst allowing the position of the crystal to be observed by microscope. Adjustments can then be made to the goniometer head settings such that the crystal is always central during rotation on the optical analyser. The goniometer head can then be transferred directly to the gonoimeter on the camera, and the crystal should now be maintained in the centre of the beam throughout data collection. Modern cameras have goniometer heads that allow adjustment of the crystal *in situ* using optic fibre technology to visualise the crystal.

Radiation sources

There are two methods by which X-rays are produced for crystallography. The first involves accelerating electrons at high voltage into a metal target (eg. a Rotating Anode X-ray source), and the second uses synchrotron radiation emitted by a Synchrotron Radiation Source - SRS. The former method is a laboratory-based technique, while the latter is limited to a few large international facilities (such as at Daresbury in Cheshire, or Grenoble in France). During the course of this project both types of source were used, but only the SRS produced useful data. Consequently only the synchrotron source will be described here, though details of the laboratory sources can be found elsewhere (Garman 1996).

An SRS uses electrons and positrons held in orbit within an enormous storage ring consisting of magnets. The electrons are accelerated around the curve of the ring and in the process they emit synchrotron radiation (a form of X-ray radiation so called because it was first observed using a synchrotron). The radiation is concentrated into a beam by specialised magnets known as 'wigglers' and directed into a shielded area called a 'hutch' where experimentation can be carried out. Personnel remain outside the hutch and so are protected from the intense beam. An SRS can provide an X-ray beam up to two orders of magnitude brighter than the best laboratory source. As crystals are exposed to X-rays they are subject to irreversible radiation damage that ultimately limits the useful data that may be collected from a crystal. The extent to which the beam initiates damage is not directly proportional to the intensity of the beam, but is a combination of factors, including X-ray intensity, exposure time, X-ray wavelength, crystal water content and the degree of scattering. The use of a high power X-ray source, such as that generated by a synchrotron, allows more useful data to be collected before a crystal is damaged. Rapidity of data collection limits exposure of the crystal, while also allowing more efficient use of the X-ray equipment. The brightness of a synchrotron beam also allows data to be collected from very small crystals (even as small as 50 μm, although at least 100 μm is desirable). The intensity and high quality optics of a synchrotron source is also ideal for collecting data from crystals with large unit cells. In recent years cryocrystallography – the freezing of crystals prior to collection, and the use of nitrogen streams to maintain crystals in the frozen state during collection - has become widespread.

The advantages of cryocrystallography are summed up in the "Notes on cryocrystallography", from the X-Ray Crystallography Group of Bernhard Rupp, in the paragraph below.

"Intrinsic limitations due to diffraction quality, can be dramatically improved by cryo-techniques, in which crystals are flash-cooled to near liquid nitrogen temperatures. Many factors contribute to improvements in data quality: Obvious benefits are reduced thermal vibrations, enhanced signal-to-noise ratio, reduced conformational disorder, and in many cases, a higher limiting resolution. An important secondary effect is the suppression of radiation damage, permitting a complete data set to be collected from one single crystal. This in turn eliminates errors from merging and scaling of data sub-sets from multiple crystals. In addition, crystal mounting is vastly simplified over conventional capillary techniques. These improvements combined lead to enhanced contrast and sharper detail in electron density maps, facilitating model building and reducing the total time required for structure determination."

Unfortunately this technique was unavailable at UCL during the period 1992 – 1994 when HIV-2 PR crystals were being subjected to data collection, and cryocrystallography systems had not yet been installed at the Daresbury SRS.

Detectors

A variety of methods are available for the detection of diffraction patterns, including the use of film. Today the image plate represents the most common form of detector, and the data obtained during this study was collected by MAR image plate.

Briefly a plate consisting of a barium halide phosphor (BaFBr) and europium (Eu^{+2}) is exposed to incident X-rays, and in the process the europium is excited into a metastable state (Eu^{+3}) at areas of the plate which were bathed in the rays. The excited portions of the plate are then stimulated to produce violet light (390 nm) by scanning the plate with a red He-Ne laser (633 nm). A photomultiplier is then able to detect the emitted light, the intensity of which is proportional to the number of absorbed X-rays. Following integration the eventual result (an output file) resembles a scanned film. The plate can be regenerated for another round of collection by exposure to bright yellow light, which erases any record of the previous image and returns the europium to its unexcited state (Eu^{+2}). Further details of image plate operation are provided by

'Modern methods for rapid X-ray diffraction data collection from crystals of macromolecules' (Garman 1996).

Data analysis

All data analysis steps were carried out using unix based CCP4 programs run on a Silicon Graphics 320. Data files for analysis were transferred to the hard disc of the Silicon graphics computer, and compressed in order to reduce the disc space they occupied. The pack_images command was used to compress image files whilst still allowing them to be read by other CCP4 programs. Files containing oscillations or stills were then visually inspected (using the ipdisp command) in order to assess whether they were suitable for further processing. If a diffraction pattern was visible then the program 'Imstills' was used, this created a file containing a list of the coordinates and intensities of the spots present on the images collected. The output files from Imstills were then fed into the autoindexing program 'Refix', which determined the crystal orientation and possible cell parameters (ie. the unit cell and space group). Initially the Refix commands were edited such that the program assumed the data related to a crystal of unknown parameters. The output was in the form of a list of potential cell parameters. These were studied to identify a space group with the highest order of symmetry, a good quality of fit (ie. < 20) and unit cell constraints that were as close as possible to the required coordinates for that particular Bravais lattice / Space group. The Refix program was then re-run having been modified to assume that the crystal has cell parameters corresponding to those just selected from the previous run's output.

Further information concerning the CCP4 programs used can be obtained from the MOSFLM users guide **mosflm_user_guide.doc** (available from Andrew Leslie, andrew@uk.ac.cam.mrc_lmb).

Further reading

If a more detailed account of any technique involved in the production of protein crystals or collection and refinement of data is required there are a numerous text books (See Appendix: Further Reading; Table 9, page 352) dealing with these topics in depth.

Chapter Three: HIV-1 Protease.

Cloning, expression, purification and crystallisation of HIV-1 protease isolated as an insolubly expressed enzyme from E. coli; and preliminary analysis of rationally designed HIV-1 PR inhibitors.

INTRODUCTION.

Expression and purification of crystallisation quality HIV-1 PR.

The aspartate proteinase of HIV-1, as discussed in chapter 1, has long been an attractive target for anti-viral drug therapy (Kramer, Schaber *et al.* 1986; Johnston, Allaudeen *et al.* 1989; Roberts, Martin *et al.* 1990). Identification, testing and refinement of active compounds necessarily requires a consistent source of the pure protease and the expression, purification and crystallisation of this enzyme has previously been described by a number of groups (Lapatto, Blundell *et al.* 1989; McKeever, Navia *et al.* 1989; Wlodawer, Miller *et al.* 1989). However, the majority of published procedures require large culture volumes, complex expression systems or numerous chromatographic procedures in order to obtain sufficient purified enzyme suitable for structural analysis and rational drug design. The development of a rapid and efficient system to overproduce and purify this viral enzyme would therefore facilitate further biochemical and biophysical studies. Our intention was to develop a tightly controlled expression system, in conjunction with an easily executed purification protocol, which together would provide HIV-1 PR of sufficient yield and quality for structural and kinetic studies.

HIV, in common with several other lentiviruses, exhibits cytotoxic effects on infected cells in culture. There is evidence implicating the viral protease as a possible cause of this cytotoxicity (Hostomsky, Appelt *et al.* 1989; Baum, Bebernitz *et al.* 1990; Korant and Rizzo 1990; Kaplan and Swanstrom 1991; Korant and Rizzo 1991). It is supposed that this is the result of proteolytic activity, since it is known that many eukaryotic proteins (especially cytoskeletal proteins) are substrates for HIV protease (Honer, Shoeman *et al.* 1991; Shoeman, Honer *et al.* 1991; Shoeman, Kesselmier *et al.* 1991; Shoeman, Mothes *et al.* 1991; Tomasselli, Howe *et al.* 1991; Tomasselli, Hui *et al.* 1991). Where HIV protease has been expressed in bacterial cells a similar effect has been noted and yields of the protease are generally low, approximately 0.3 mg

protease per litre of cell culture (0.75 – 2.5% of total cell protein) (Danley, Geoghegan *et al.* 1989). In terms of practicality protein yields in excess of 1 mg per litre of cell culture are preferred prior to engaging in crystallographic analyses. Some studies have therefore focused on synthetically produced enzyme (Wlodawer, Miller *et al.* 1989), but this is an expensive option. The expression system described here produced the HIV-1 protease in an insoluble form, and with a slight improvement in yield over previous protease expression protocols. The system provided sufficient protease for use as a control in assays of retroviral protease activity, for the testing of protease inhibitors, and for use in crystallisation trials.

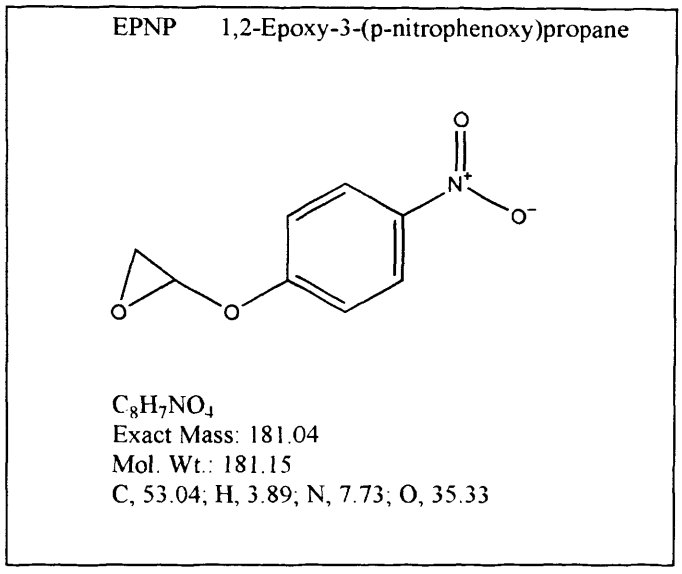
A two-step chromatographic purification procedure was developed, producing HIV-1 PR that was > 95% pure as determined by SDS-PAGE, which was enzymically active, and crystallised readily under conditions described for the solubly expressed protein (Lapatto, Blundell *et al.* 1989; McKeever, Navia *et al.* 1989). Crystallisation of HIV-1 protease refolded from insoluble bacterial inclusion bodies had not been previously reported.

Preliminary analysis of novel, rationally designed, HIV-1 PR inhibitors.

The development of compounds able to inhibit HIV proteases *in vivo*, preventing viral maturation and consequently limiting the progression of AIDS in man is the ultimate aim of work in this field. This chapter also describes the initial stages in development of two potential inhibitor compounds.

Early inhibitors were almost exclusively based on pepstatin and peptide substrate analogues. However, these are reversible, competitive inhibitor compounds and thus do not permanently disable the target enzyme since binding is non-covalent. Consequently viral strains with mutant proteases resistant to such inhibitors have arisen. Inhibitors able to bind covalently with the enzyme active site would produce permanent protease inhibition. Such compounds would also be less susceptible to the development of resistant strains since they target the conserved active site residues. Irreversible 'suicide' compounds therefore have a clear advantage over conventional inhibitors. The two compounds described here were based on this strategy, and were synthesised at UCL by Dr. George Nicolau. The principle of irreversible HIV PR inhibition has been demonstrated using 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) (Meek, Dayton *et al.* 1989), which is a weak and non-specific irreversible inhibitor of aspartyl proteases (fig. 1), reacting with the catalytic aspartate residues of pepsin (Tang 1971) and penicillopepsin (James *et al.* 1977). It was subsequently shown to inhibit HIV-1 PR in a pH dependant manner, suggesting modification of a single active site aspartate (Meek, Dayton *et al.* 1989).

Figure 1 Structure of EPNP (1, 2-Epoxy-3-(p-nitrophenoxy) propane).



An epoxide moiety was incorporated into the first of the inhibitor compounds described here, and an aziridine moiety into the other, each with the intention of provoking a covalent interaction between the compound and a catalytic residue of HIV-1 PR.

The aziridine moiety, like epoxide, is a highly reactive group consisting of a strained ring – with the exception that the aziridine ring contains Nitrogen rather than Oxygen. Both epoxides and aziridines are susceptible to nucleophilic ring opening, thus it was anticipated that they might both be able to inactivate HIV PR by alkylation of an active site aspartate. Both inhibitors were based on the following structure:

BOC-Phe – active moiety – O-methyl where BOC = *tert*-Butoxycarbonyl

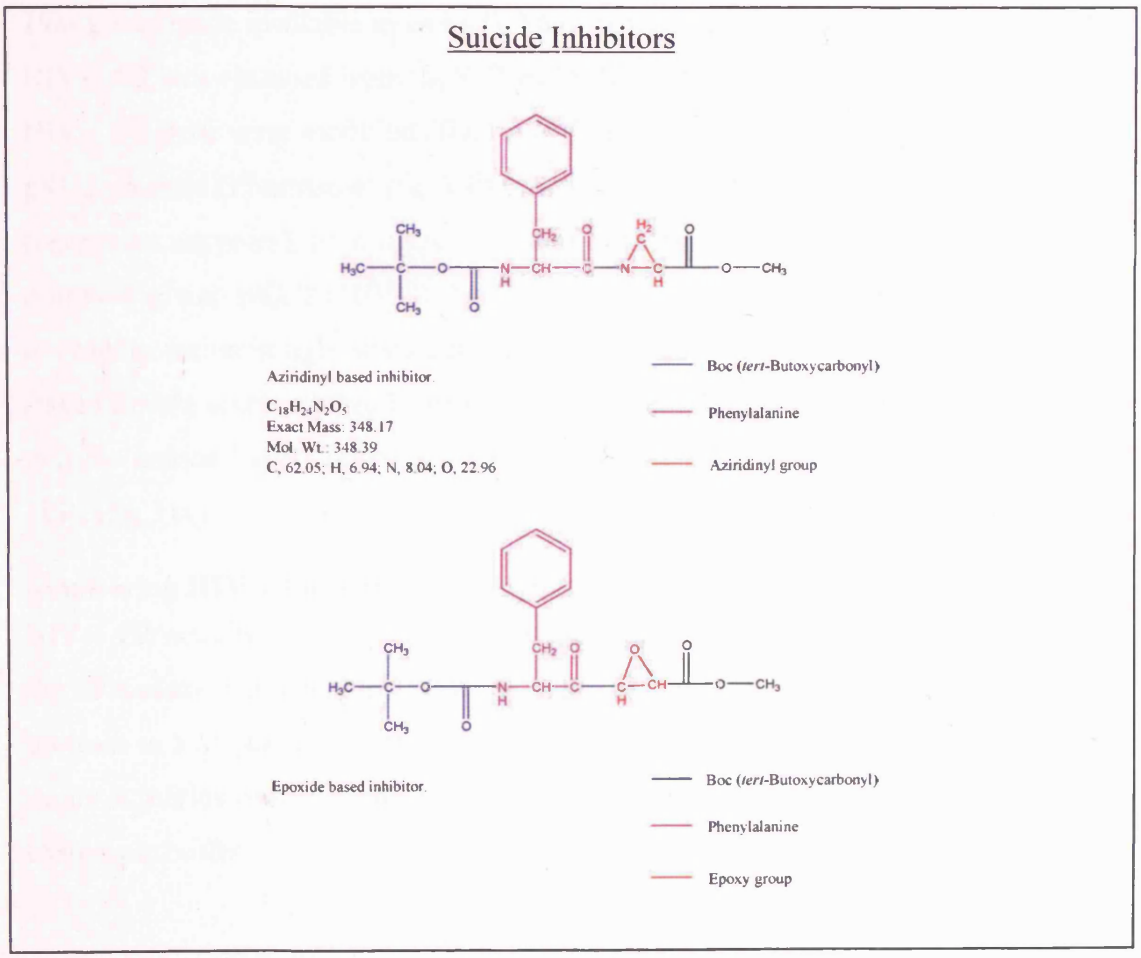
(see fig. 2).

The Phenylalanine was intended to engage with the substrate-binding cleft at S1 and thus locate the reactive group in close proximity to the catalytic residues.

HIV-1 PR produced by means of the expression and purification method described in this chapter was used in conjunction with a chromophoric assay (Richards, Phylip *et al.* 1990) to test the efficacy of these inhibitors. Preliminary results are presented here demonstrating that the aziridine based compound was able to inhibit HIV-1 PR activity. Further experiments to confirm the mode of action and allow kinetic

characterisation would have been desirable and are therefore briefly discussed. However, a limited and finite supply of the compounds precluded them.

Figure 2 Compounds intended to irreversibly inhibit HIV-1 PR.



MATERIALS AND METHODS.

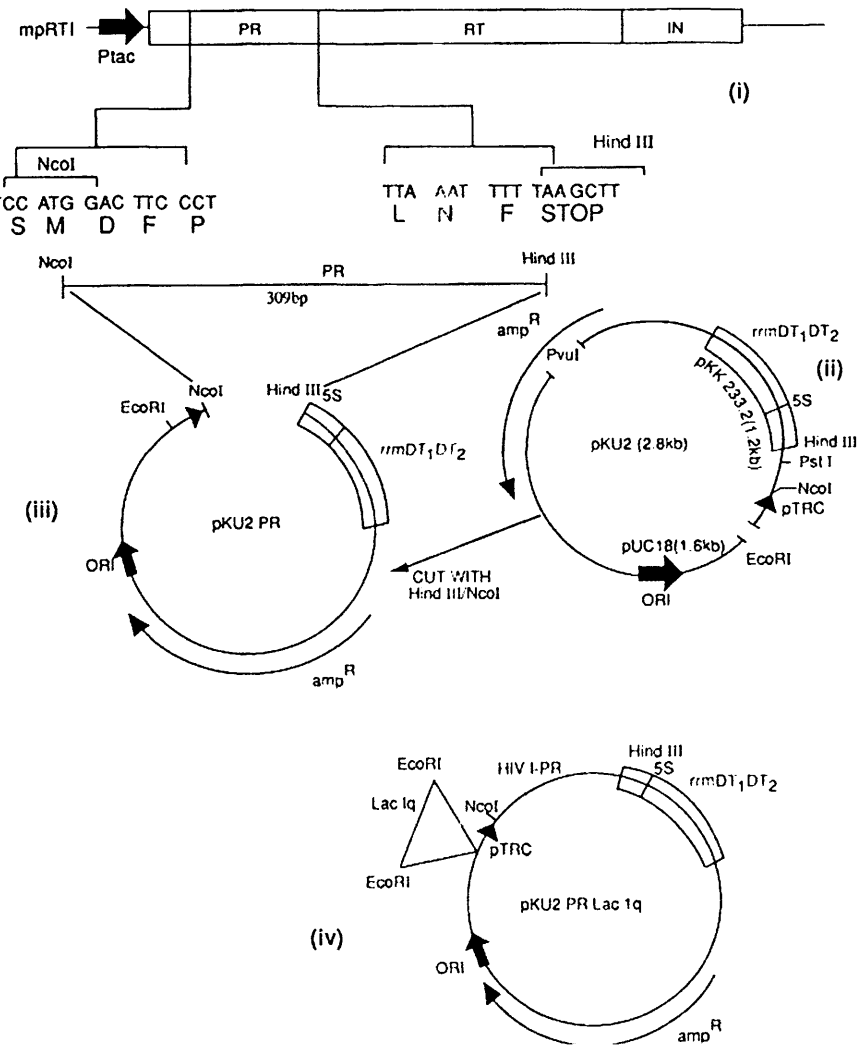
Construction of the HIV-1 PR expression vector pKUlacIq (HIV-1PR).

A Reverse Transcriptase expression vector, mpRT1, (Larder, Purifoy *et al.* 1987) containing the pol operon derived from the HIV-1 strain HXB-2 (Shaw, Hahn *et al.* 1984), was made available to us by G. Darby of Wellcome. The 309 bp gene encoding HIV-1 PR was obtained from mpRT1 by PCR. In the process the 5' and 3' ends of the HIV-1 PR gene were modified (fig.3 i) to facilitate cloning into the pKU2 vector. The pKU2 plasmid (Pharmacia) (fig.3 ii) and the HIV-1 PR gene were each digested with (restriction enzymes), then ligated to insert the HIV-1 PR gene downstream of the trc promoter giving pKU2 (HIV-1PR) (fig.3 iii). The trc promoter is IPTG inducible and in order to maintain tight transcription regulation of the HIV-1 PR gene the lacIq gene (taken from a vector - pMC9 - provided by Richard Skinner, Wellcome) was inserted into the unique EcoRI restriction site of pKU2 (HIV-1PR) to yield $pKU2_{lacIq}$ (HIV-1PR) (fig.3 iv).

Monitoring HIV-1 PR activity by SPA

HIV-1 PR activity was monitored throughout the purification procedure by means of the SPA assay, described in Chapter 2. It should be noted that when an aliquot of the protease in 5 M guanidine (or 8 M urea) is pipetted into a larger volume of SPA assay buffer it refolds instantly thus allowing detection of active fractions whilst utilising a chaotropic buffer.

Figure 3 Cloning strategy for the expression of HIV-1 PR.



Key to cloning strategy.

(i) Site directed mutagenesis of mpRT1 to introduce an in frame ATG and NcoI site at the 5' end and a TAA translational stop and HindIII site at the 3' end of the protease gene.

(ii) Construction of the expression vector pKU2 using the 1.2 Kb PvuI/EcoRI fragment from a vector PKK233-2 which contains the pTrc promoter and the rrmDT1DT2 termination signals and the 1.6Kb PvuI/EcoRI fragment from the vector pUC18, which contain the pUC origin of replication.

(iii) Insertion of the 309bp NcoI/HindIII protease fragment into pKU2 down stream of the ptrc promoter.

(iv) Insertion of the lac Iq repressor gene into the unique EcoRI restriction site of pKU2PR to form pKU2PR lac Iq.

Large scale cell growth and protein expression

The pKUlacIq (HIV-1PR) plasmid was used to transform *E.coli* strain MC1061, and grown as described in Chapter 2 (for transformation see page 138, and for cell growth see page 145). Optimum yields were obtained when the culture was grown to late log phase (OD_{600} 0.7), induced with 2 mM IPTG, and grown for a further two hours prior to harvesting. A litre culture typically yielded 2.5 g of cell paste. Protease expression was monitored both by Western blotting and SPA. Both techniques are described in Chapter 2 (see page 153).

Extraction of insoluble HIV-1PR

Frozen cell pellets (~20 g) were thawed and the *E.coli* lysed using a probe sonicator (Soniprep 100). The cell suspension was chilled on ice throughout the lysis procedure, which consisted of 10 X 30 s bursts of the sonicator with a 60 s pause between bursts. The lysate was clarified using a Beckman L8-M ultracentrifuge in conjunction with an SW-28 rotor at 28000 rpm for 30 min at 4°C. The supernatant was discarded, but the pellet was resuspended in 30 ml of buffer A, resonicated and re-centrifuged, as above.

Table 1 Buffers A and B.

Buffer A:
50 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT
Buffer B:
5 M Guanidine HCl, 100 mM MES pH 6.5, 1 mM EDTA, 5 mM DTT

The supernatent was again discarded, and the pellet was solubilised at room temperature by resuspension in 30 ml of Buffer B, followed by agitation overnight.

Insoluble particulates were removed by ultracentrifugation (as above), this time for 90 minutes at 28000 rpm and 4°C. The pellet was discarded and the clear, golden coloured supernatent was retained. The volume of the protein solution by this stage was ~50 ml. The solution could be stored at 4°C for several weeks without any apparent effect on subsequent HIV-1 PR activity, though usually chromatography immediately followed extraction.

Size Exclusion Chromatography of insoluble HIV-1 PR

Following extraction and dissolution, the insolubly expressed proteins were subjected to size exclusion chromatography, using a 5 x 90 cm Sephacryl S-200 HR column equilibrated with Buffer B. The sample was introduced to the column via a 50 ml superloop then a single column volume of Buffer B was pumped through the column at 1 ml min^{-1}. Once the void volume of 120 mls had eluted, 20 ml fractions were collected for analysis. Those fractions identified as having HIV-1 protease activity, using the SPA, were pooled; giving approximately 160 ml of proteolytically active sample (see figs. 4 and 5 pages 192-193, and Appendix tables 1 and 2, pages 343-344).

Refolding the semi-purified HIV-1 PR

The 160 ml active fraction was adjusted to 10% (v/v) glycerol and the protease refolded by a two step dialysis procedure at 4°C. The first dialysis step was against buffer C over an 8 hour period with four changes of buffer, the second was against Buffer D and carried out for 16 hours at 4°C.

Table 2 Buffers C and D.

Buffer C:
100 mM MES pH 6.5, 2 M Guanidine HCl, 1 mM EDTA, 5 mM DTT, 10% glycerol
Buffer D:
50 mM MES pH 6.5, 1 mM EDTA, 5 mM DTT, 10% glycerol

A white precipitate was formed during dialysis, which was removed by ultracentrifugation (as described, at 25000 rpm and 4°C for 90 mins). The supernatant was retained for further chromatography.

Ion Exchange Chromatography of refolded HIV-1 PR

The clarified solution of refolded HIV-1 PR was loaded on to an S-Sepharose high load 16/10 ion exchange column (at 1 ml min^{-1}), previously equilibrated with Buffer D. The column was washed with 5 column volumes of Buffer D, or until a steady baseline was reached. Protein was eluted in Buffer D by means of a salt gradient, initially rising from 0 M to 200 mM NaCl at a rate of 100 mM per column volume. Then, after a period of 2 column volumes at 200 mM NaCl, the gradient was continued up to 1 M NaCl. Protein solution eluted during the 200 mM NaCl wash

was collected, subjected to SPA analysis and shown to contain HIV-1 PR activity (see results fig. 6). Purity was assessed by means of coomassie stained SDS-PAGE.

Concentration and storage of purified HIV-1 PR

Active fractions from the S-Sepharose column were pooled (15 ml total volume) and dialysed against Buffer E.

Table 3 Buffer E.

10 mM MES pH 6.5, 1 mM DTT, 1 mM EDTA

This removed unwanted NaCl and glycerol (which may have interfered with crystallography), while also reducing the concentration of MES. The volume of solution was reduced using a Centricon 10 (as described in Chapter 2, pages 161-162) until a protein concentration of approximately 5–6 $mgml^{-1}$ was reached (as determined by UV spectroscopy at λ = 280 nm and given an extinction coefficient for HIV-1 PR native dimer of 25800 $cm^{-1}M^{-1}$). The concentrated solution was stored in aliquots of 50 µl at -70°C pending further use.

Determination of kinetic parameters for the purified HIV-1 PR

The procedure for determination of K_M, K_{cat} and V_{max} is described in Chapter 2 (pages 164-165). The reaction buffer used for kinetic analysis of HIV-1 PR was 100 mM Sodium Acetate pH 5. A 10 mM stock of chromogenic substrate III from BACHEM (Fig. 7 b.) was freshly prepared in water, and reactions were carried out over a range of final substrate concentrations from 0 – 200 μM. Reactions were repeated in triplicate and the mean initial velocity calculated at each substrate concentration. HIV-1 PR concentration was constant throughout at 0.6 μgml^{-1} final. Values for K_M K_{Cat}, and V_{max} were determined using the Enzfitter program.

Crystallographic analysis of HIV-1 PR

Crystals were grown by the hanging drop method (Chapter 2, page 170) under conditions described by McKeever *et.al.* (McKeever, Navia *et al.* 1989). Briefly, a 2 μl drop consisting of 50% protein solution and 50% crystallisation solution was incubated at 4°C over a well containing 1 ml of the crystallisation solution. The crystallisation solution consisted of 100 mM imidazole / HCl pH 7.0, 250 mM NaCl, 10 mM DTT and 1 mM NaN_3. Typically crystals were seen after 3-4 days. Data collection was carried out at the Synchrotron Radiation Source, Daresbury, and processed as described in Chapter 2 (pages 174-178).

Computer aided design of HIV-1 PR inhibitors

The inhibitors were designed with the aid of three dimensional (3D) computer models, based on the available HIV-1 protease structural data in early 1992 (Protein Databank files 4HVP, 5HVP, 7HVP, 8HVP, 9HVP, 3PHV). Laurence Pearl and George Nicolau carried out the modelling and inhibitor at UCL, using the modelling programs Quanta and CHARMm (available from Accelrys Ltd.).

Chemical synthesis of the HIV-1 PR inhibitors

The inhibitors were synthesised by George Nicolau at UCL and characterised by the UCL Chemistry department using Mass Spectrometry (MS), and Nuclear Magnetic Resonance (NMR) to ensure that they had been synthesised correctly (George Nicolau, personal communication).

Assays to detect inhibition of HIV-1 PR activity

Assays using the chromogenic substrate III, from BACHEM, (Richards, Phylip *et al.* 1990) were carried out essentially as described in Chapter 2 (pages 164-165) to determine the extent of HIV-1 PR activity in the presence and absence of the various potential inhibitors, using Acetyl-Pepstatin (from BACHEM) as the control inhibitor. The reaction buffer was identical to that used for the K_M determination described above, and total reaction volume was 1ml. Substrate III and HIV-1 PR were used at final concentrations of 50 μM and 0.6 μgml^{-1} (27.9 nM) respectively. The inhibitor compounds and the acetyl-pepstatin control were dissolved in DMSO and tested at a final concentration of 1 μM. The dual beam spectrophotometer used (see Chapter 2, page 164) allowed the use of a blank containing the inhibitor compound under test, thus controlling for the absorbance due to these compounds.

Dialysis of suicide inhibitor – protease complex

Microdialysis (as described in Chapter 2, page 158) of reaction mixtures containing HIV-1 PR alone (control mix) or HIV-1 PR plus inhibitor compound (inhibitor mix) was carried out to determine whether the suicide compounds were able to irreversibly inhibit the protease. A non-covalently bound inhibitor would be able to dissociate from the active site, and diffuse out of the reaction mixture into the surrounding dialysis buffer. A compound diffusing into the dialysis buffer would be massively diluted, thus its return to the reaction mixture is highly improbable. Therefore if a compound were not a true suicide inhibitor, dialysis should allow restoration of protease activity. The control reaction mixture was 3 ml final volume and consisted of HIV-1 PR reaction buffer supplemented with final concentrations of 40% v/v glycerol, 1 mM DTT and 10 % v/v DMSO, plus HIV-1 PR at 0.6 μgml^{-1} (27.9 nM) final concentration. The presence of glycerol and DTT was intended to stabilise the enzyme and preserve activity during the dialysis procedure, whilst the DMSO was required to maintain solubility of the inhibitors. The inhibitor mixtures were made up in the same way (taking into account the additional DMSO added with the inhibitor solutions) and contained their respective inhibitor compound at a final concentration of 100 μM. Dialysis was carried out overnight, with gentle stirring, at 4°C in 250 ml of HIV-1 PR reaction buffer also supplemented with 40% v/v glycerol, 1mM DTT, and 10% v/v DMSO.

RESULTS AND DISCUSSION

Protein expression

The majority of expression systems reported for over production of the HIV-1 PR were developed as compromise solutions, in order to overcome the toxicity of the protease gene product in the *E. coli* host and the need to obtain sufficient enzyme for crystallisation (Danley, Geoghegan *et al.* 1989; McKeever, Navia *et al.* 1989; Goobar, Danielson *et al.* 1991). Some of these expression systems utilise very careful control of growth conditions (Danley, Geoghegan *et al.* 1989) in order to keep the protease gene repressed, or complex dual vector systems for expression (Cheng, McGowan *et al.* 1990), and often require large volumes of bacterial culture, to obtain sufficient enzyme for purification (McKeever, Navia *et al.* 1989).

The pKU2 vector, which formed the basis of our construct, was originally designed to be a high copy number vector (having a pUC origin of replication) with a strong inducible promoter (trc promoter) in order to produce large amounts of expressed protein. However, expression of the HIV-1 PR gene from pKU2 PR in *E.coli* resulted in poor bacterial growth. This could be explained by previous observations that the protease may be toxic to the bacterial host (Baum, Bebernitz *et al.* 1990). It was therefore desirable to ensure tight control of expression by introduction of the Lac Iq repressor gene. No improvement was seen even when the plasmid was used to transform the *E.coli* host TG1, which carries a single copy of the Lac Iq gene. Cloning of the Lac Iq gene into pKU2 PR (fig 3) to yield $pKU2_{laclq}$ (HIV-1 PR) greatly improved the stability of the clone, and allowed it to be maintained adequately in the host *E.coli* strain MC1061. Indeed the growth characteristics exhibited by the *E.coli* transformed with pKU2 PR $_{(Laclq)}$ (HIV-1PR) were similar to that of *E.coli* harbouring the parent vector pKU2 (HIV-1PR) alone. The rationale for placing the Lac Iq gene on the same expression vector as the protease gene was to maintain equal copy numbers of the repressor protein and its target promoter. Consequently the repressor protein is not titrated out by the high copy number of the promoter sequence due to the multicopy nature of the pKU2 plasmid. A commercial vector, pTrc99A (Amersham Pharmacia Biotech), has since become available which shares these characteristics.

The strength of the trc promoter from pKU2 resulted in the accumulation of significantly more enzyme in the insoluble fraction than we had previously observed

using other systems. Since better yields of the protease could be produced as insoluble material it was decided to exploit this by extracting, purifying and refolding protease from the insoluble fraction of the cell lysate.

Purification of HIV-1 PR

The protease was purified by molecular sieving through Sephacryl S-200, following solubilisation of the pellet fraction with guanidine. Protease activity was detected by SPA in fractions 8 to 15 (fig 4) of the low molecular weight peak eluting from the column (fig 5). The SPA data is presented in tables 1 and 2 of the appendix. Refolding by dialysis and subsequent analysis by PAGE of the proteins contained within this peak fraction showed that it contained a number of proteins of a MW range from ~ 6 - 20 kD (fig. 7). The HIV-1 protease contained within this fraction was estimated to be > 60% pure.

Figure 4 SPA data plot.

Fractions exhibiting ≥ 70% relative activity (8 – 15 inclusive) were pooled.

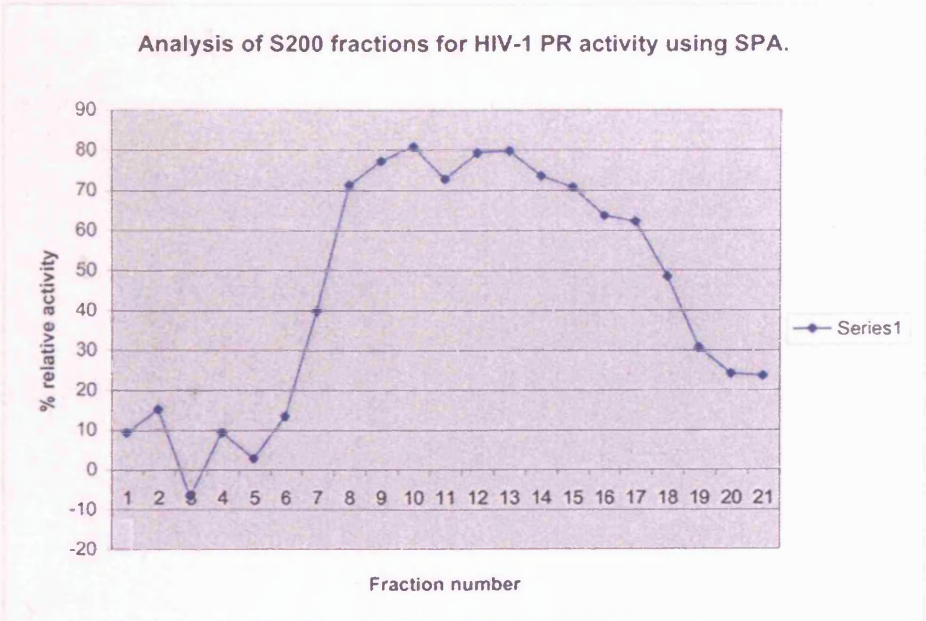
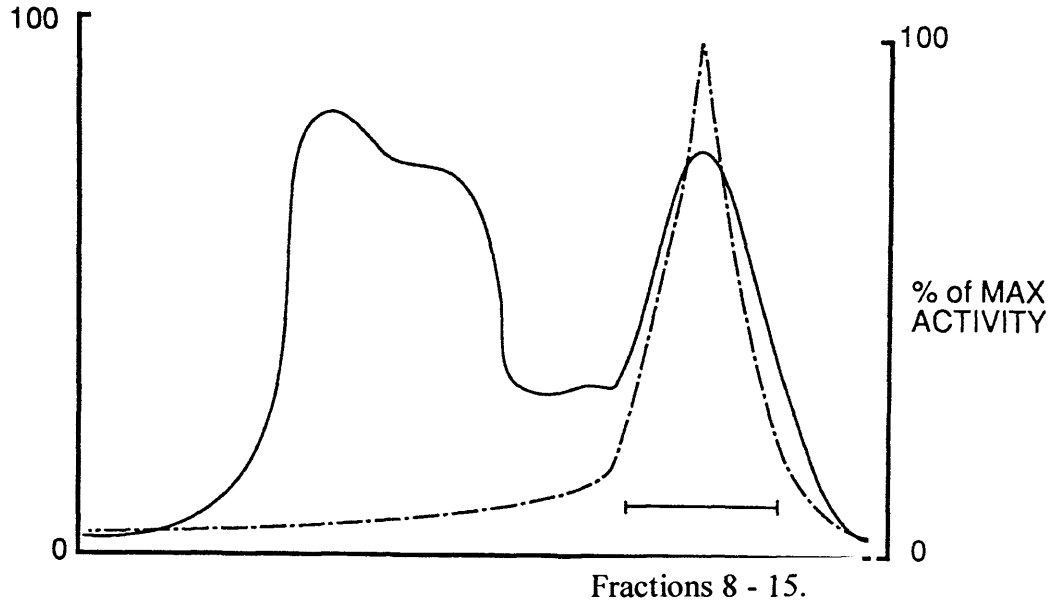


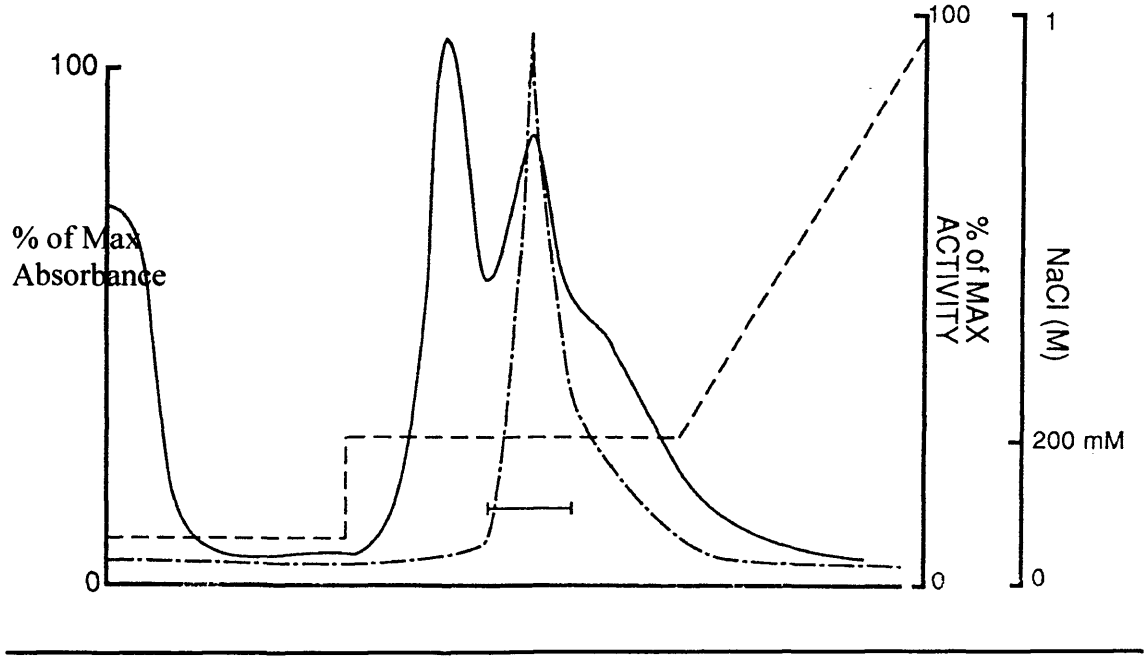
Figure 5 Size Exclusion Chromatography of insoluble HIV-PR.



The pellet following cell lysis was dissolved in 5M Guanidine HCl, and loaded onto a Sephacryl S200 size exclusion column. This figure shows the resulting chromatogram, the solid line represents absorbance (mAu) at λ = 280 nm, the broken line represents HIV-1 PR activity. Most protease activity was contained within the low molecular weight peak. The bar represents fractions 8 to 15, which were pooled and retained.

The refolded protease was further purified by S-sepharose chromatography. This resulted in the elution of two protein peaks (fig. 17), the protease being isolated from the second peak. A simple salt gradient failed to separate these two conjoined peaks of protein adequately. However, by holding the NaCl gradient at 200 mM for 2 column volumes an improved separation was achieved and highly purified active HIV-1 protease obtained.

Figure 6 Cation Exchange Chromatography of refolded HIV-1 PR.



Refolded HIV-1 PR was loaded onto a S-Sepharose Hi-load column, and eluted by an NaCl stepped gradient. Protease activity (detected by SPA) was found exclusively in the second peak eluting from the column during the 200 mM NaCl step.

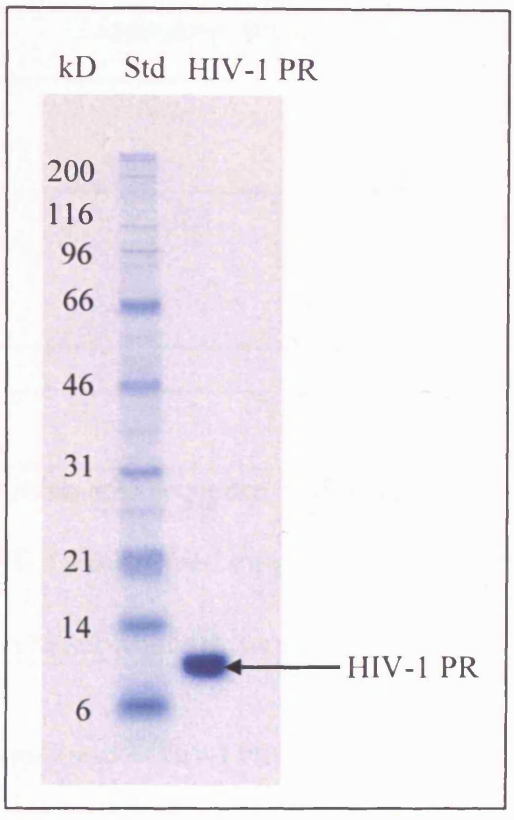
Complete line: absorbance at $\lambda = 280$ nm

Dashed and dotted line: % protease activity

Dashed line: [NaCl] M.

The final HIV-1 PR yield was typically between 0.5 – 1.0 mg of pure enzyme from each litre of bacterial culture harvested, and was essentially 90% pure as assessed by SDS-PAGE analysis (fig. 7).

Figure 7 SDS PAGE analysis of the purified HIV-1 PR.



Characterisation of the purified HIV-1 PR

The SPA assays used to track activity throughout purification had already indicated the presence of HIV-1 PR. N-terminal analysis and Mass Spectrometry (both carried out a Wellcome Laboratories, Beckenham) established that the purified protein was indeed HIV-1 PR (data not shown).

K_{Cat} and K_M determination of HIV-1 PR

The SPA substrate was unsuitable for a K_M determination since the SPA substrate cleavage site does not correspond to a natural cleavage sequence, and because it is not possible to estimate the concentration of peptide attached to each microsphere. In addition to this our scintillation counters were unsuitable for the measurement of initial velocity.

A K_M for the purified protease enzyme of 27 μM and a K_{Cat} of 0.0657 sec^{-1} was obtained using the chromogenic substrate (fig. 8 a and b), which compared well with the value obtained by Pennington *et al.* (Pennington, Dick *et al.* 1990).

Figure 8 The hydrolysis of chromogenic substrate by purified HIV-1 PR.

a. Lineweaver-Burke transformation.

Variable	Value
K_{Cat}	0.0657 sec^{-1}
K_M	27.8 uM

b. Amino acid sequence of chromogenic substrate.

H-His-Lys-Ala-Arg-Val-Leu-p-nitro-Phe-Glu-Ala-Nle-Ser-NH_2

HIV Protease substrate III, suplied by Bachem, product no. H-9035

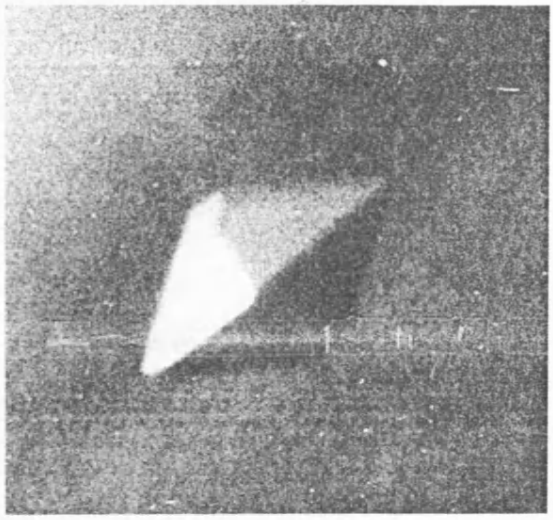
Values of K_M and K_{cat} were established for HIV-1 PR activity by means of chromogenic assay.

Crystallisation

Tetragonal bi-pyramidal crystals appeared after four days. These crystals were approximately 100 μm in length and produced measurable diffraction intensities to a resolution of 2.9 Å, having the same unit cell dimensions as previously described (McKeever, Navia *et al.* 1989; Wlodawer, Miller *et al.* 1989) (fig. 9).

The ability to grow these crystals confirmed that the enzyme obtained by the procedures described was of suitable quality for structural studies. Since the unit cell was the same as that reported by groups using the solubly expressed enzyme it can be inferred that our protein was correctly refolded and in its native configuration.

Figure 9 Tetragonal bipyramidal crystal of HIV-1 protease.



Crystals were grown to approximately 100 μm in length, were tetragonal and the unit cell dimensions were a=b=50.3 Å and c=106.8 Å, $\alpha = \beta = \gamma = 90°$, with a space group of either $P4_12_12$ or $P4_32_12$.

Inhibitor testing.

The SI compounds were intended to covalently bind the active site (via an epoxide or aziridine moiety respectively), permanently inactivating the protease (figs. 10 and 11).

Figure 10 The anticipated mode of action for the epoxide compound.

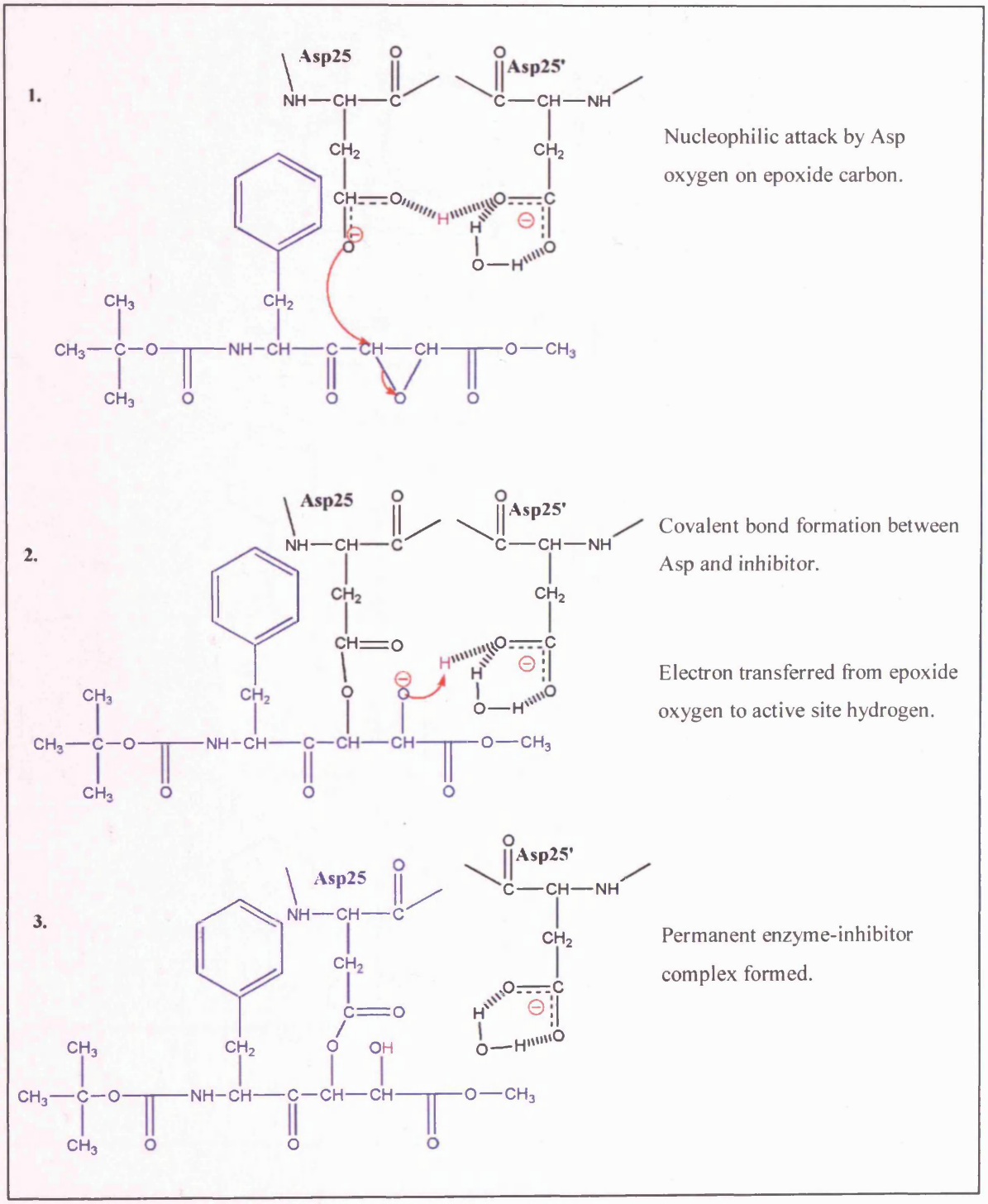
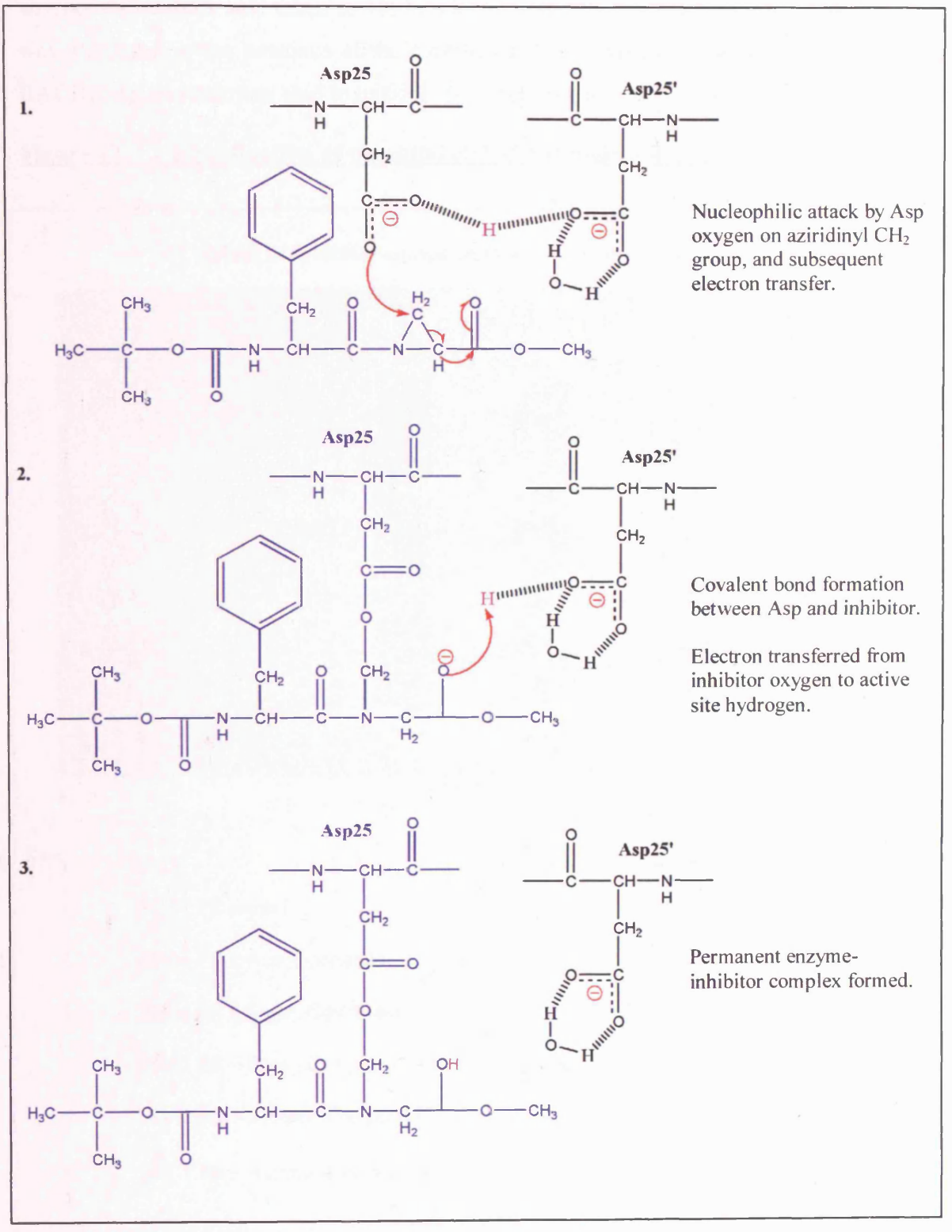
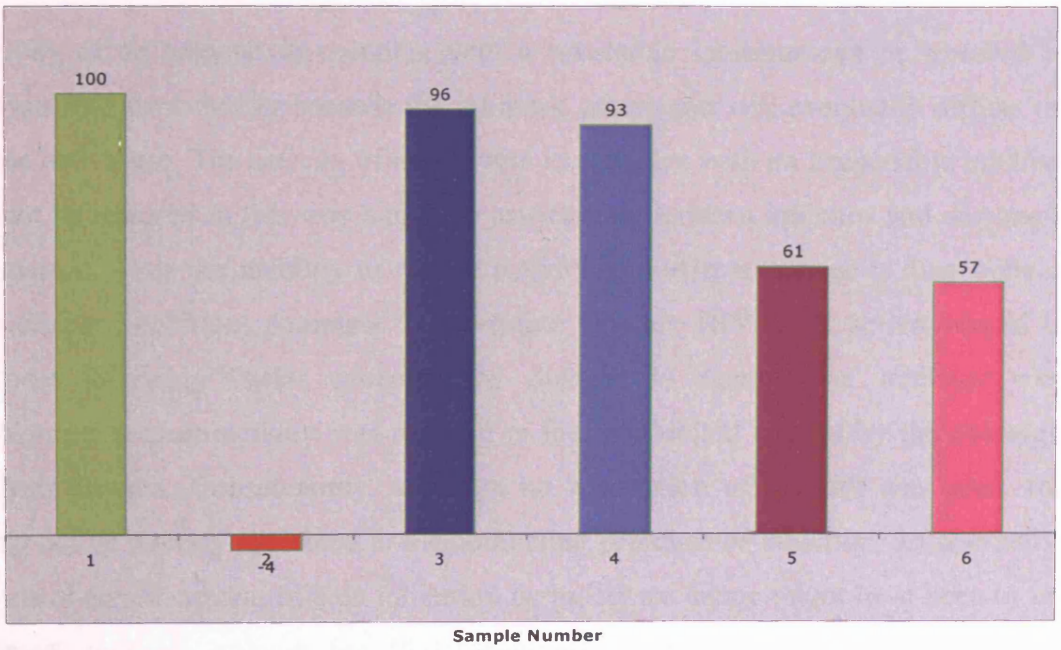


Figure 11 The anticipated mode of action for the aziridine compound.



The initial activity rate for HIV-1 PR was measured by chromogenic substrate in the absence of inhibitor and taken as 100% relative activity, or 0% inhibition. No activity was measured in the presence of the control inhibitor, Acetyl-Pepstatin (supplied by BACHEM), and this was said to exhibit 100% relative inhibition (fig. 12).

Figure 12 Investigation of potential HIV-1 PR inhibitor compounds.



Key:

1	HIV-1 PR control
2	HIV-1 PR + Acetyl-pepstatin
3	HIV-1 PR + Epoxy Compound
4	HIV-1 PR + Epoxy Compound (2 hrs)
5	HIV-1 PR + AziridineCompound
6	HIV-1 PR + AziridineCompound (2 hrs)

Absorbance was measured at 600 nm for 200 seconds and the rate of change calculated for each sample. Reaction rate (Au/s) for the HIV-1 PR control was taken to be 100%, and the rates for all other samples were expressed as percentages relative to the HIV-1 PR control.

The epoxide compound did not appear to inhibit the protease at all, and even after two hours incubation with this compound the protease retained virtually 100% relative activity (fig. 12). Whilst the reason for this is unclear, it may indicate that the epoxide moiety could not adopt a suitable position relative to the catalytic residues while the compound occupied the protease active site. In contrast, the aziridine compound reduced relative protease activity by 40% (fig. 12). Incubation of inhibitor and enzyme for two hours prior to assay also resulted in a 40% reduction in relative activity.

Activity of an enzyme in complex with a reversible inhibitor can be restored by dialysis into fresh buffer because the inhibitor compound will eventually diffuse out of the active site. The activity of an enzyme in complex with an irreversible inhibitor cannot be restored in this way since the association between inhibitor and enzyme is permanent. Thus the inability to restore activity by buffer exchange is diagnostic of irreversible inhibition. Attempts to determine whether HIV-1 PR activity could be restored following buffer exchange by dialysis to remove the inhibitor were ambiguous because activity was reduced in the uninhibited control by the overnight dialysis process. Consequently, although no restoration of activity was seen, this could not be directly attributed to the continuing presence of inhibitor. An alternative means of demonstrating suicide inhibition by buffer exchange might have been to use a small desalting column (eg PD10 column). This would have enabled buffer exchange (and thus removal of all unbound inhibitor) to be achieved much more quickly, and thus eliminated the problem of natural deterioration in protease activity with time seen in the dialysis control. However, it might not provide sufficient time for dissociation to occur. Mass spectroscopy could have provided a more conclusive method for demonstrating that the compound was irreversibly inhibiting the protease. Where the compound was covalently bound to the protease an increase in the mass of the protease consistent with the mass of the inhibitor compound would have been detectable. A similar non-covalent interaction would not have generated this effect since a non-covalent interaction would be extremely unlikely to survive the ionisation process involved in conventional mass spectrometry. Furthermore, enzymatic digest of the inhibitor-protease complex (using Trypsin) and subsequent analysis of the resultant peptides by tandem MS methods could have demonstrated the exact site of covalent interaction. Ultimately a crystal structure of the protease in complex with the

compound would have provided a confirmation of covalent interaction, and the opportunity to improve the compound design.

Although preliminary, these results suggest that the aziridine compound is a weak inhibitor of HIV-1 PR and may therefore be a useful lead in the further development of a suicide inhibitor. A more detailed analysis of this compound is required to ensure that it is inhibiting the protease by the intended mechanism, and kinetic experiments would be desirable to characterise the inhibition further. It would also be interesting to investigate the effect these compounds, on HIV-2 and SIV_{AGM} protease activity. However, further experiments on these compounds were precluded by their finite supply.

Subsequent to the work described in this Chapter, crystallographic analysis has demonstrated that EPNP is able to bind covalently to a single catalytic aspartate residue of SIV protease, thus confirming its mode of action (Rose, Rose *et al.* 1993), and other studies have demonstrated that epoxide based compounds can be effective as irreversible HIV-1 PR inhibitors (Salto, Babé *et al.* 1994).

SUMMARY.

An *E.coli* expression system producing insoluble HIV-1PR was developed which was able to produce more mg of active pure protease per litre of culture than had previously been possible. The expression and purification procedures described in this chapter were simpler than contemporaneously published HIV-1 PR purification procedures (Danley, Geoghegan *et al.* 1989) and represented an improvement in terms of yield and ease of protein production over other published methods (Danley, Geoghegan *et al.* 1989; Darke, Leu, *et al.* 1989).

The system gave protease yields that exceeded those of soluble expression systems, presumably due to an absence, or reduction of cytotoxicity exerted by the insoluble protease. Furthermore the insoluble protease simplified purification, allowing a two stage chromatographic procedure to suffice; and utilised widely applicable chromatographic materials, rather than expensive and highly specific affinity resins. This represented an improvement on many of the existing methods. The procedure might also be adopted for other proteins exhibiting toxicity in their soluble forms, providing they can be successfully refolded.

A yield in terms of HIV-1 PR per volume of cell culture of 0.5-1.0 mg/L was typically obtained, which compares favourably with many other reports of retroviral protease expression (Danley *et al.* 1989, Strickler *et al.* 1989, Wan and Loh 1995). However, it is noted that expression was not as high as is achieved with certain other proteins using similar expression systems. This suggests either that there was still significant cytotoxicity, or that other factors may have been limiting protease expression. Similar observations have been made where the protein of interest exhibited poor codon usage (Graham, Atkinson *et al.* 1993; Martin, Vrhovski *et al.* 1995), and this may also be a factor here. Codon usage for HIV-1 PR was compared with that of genes encoding *E.coli* proteins in general and the result showed that the HIV-1 PR gene contains a high number of codons which are used only rarely, if at all, by *E.coli*. The view that codon usage may be a limiting factor when expressing HIV-1 PR in *E.coli* is supported by the work of Rangwala *et al.* (Rangwala, Finn *et al.* 1992), where expression levels improved following synthesis of an HIV-1 PR coding sequence optimised in terms of codon usage for expression in *E.coli*. In light of this our system

could have been similarly optimised, perhaps by utilising the Recursive PCR method later developed in our laboratory (Prodromou and Pearl 1992)[10].

The protease was found to have a K_M = 27.8 μM and a K_{Cat} = 0.0657 sec^{-1} - values consistent with those reported by others using the same substrate (Pennington, Dick *et al.* 1990).

Crystals of the apoenzyme diffracted x-rays to below 2.9 Å resolution and had the same unit cell as those produced by other groups (McKeever, Navia *et al.* 1989) using either solubly expressed or synthetically produced protein (Wlodawer, Miller *et al.* 1989).

A preliminary evaluation was carried out on two rationally designed compounds intended to inhibit the HIV-1 PR by means of covalent binding to the active site. One of these, an aziridine compound designed to bind the active site irreversibly, was shown to inhibit HIV-1 PR activity, and further studies were discussed.

HIV-2 poses as serious a threat human health as HIV-1, consequently compounds able to inhibit HIV-2 PR are also required. The expression of HIV-2 PR was known to be as problematic as that of HIV-1 PR, meanwhile much less work had been carried out on HIV-2 PR and no structural information was yet available. It was therefore decided to apply some of the information gained from working on HIV-1 PR to the study of HIV-2 PR with a view to carrying out structural studies on the enzyme. The following chapter (Chapter 4) describes these efforts.

ACKNOWLEDGEMENTS

Thanks to Richard Skinner for the plasmid pMC9 from which the lac 1q gene was obtained, John Champness and Patrick Bryant for collecting X-ray crystallographic data.

[10] This option is further considered with respect to SIV_{AGM} PR in Chapter 7.

Chapter Four: HIV-2 Protease.

The purification and crystallisation of recombinant HIV type 2 protease.

INTRODUCTION.

Whilst both Serotypes of HIV cause disease in man (Essex and Kanki 1988; Marlink, Ricard *et al.* 1988), the majority of studies, at the time of this work, had concentrated on HIV-1 PR (Lapatto, Blundell *et al.* 1989; McKeever, Navia *et al.* 1989; Navia, Fitzgerald *et al.* 1989; Wlodawer, Miller *et al.* 1989). However, the differences in substrate specificity (Le Grice, Ette *et al.* 1989; Wu, Carr *et al.* 1990), inhibitor binding (Richards, Broadhurst *et al.* 1989) and amino acid sequence (Guyader, Emerman *et al.* 1987) between the HIV-1 and HIV-2 proteases were sufficient to warrant studies of HIV-2 PR. Consequently it was desirable to obtain a crystallographic structure of HIV-2 PR, in the pursuit of a more complete understanding of the HIV proteases generally. At the inception of this project a structure for HIV-2 PR was not available, and the preliminary crystallisation of native HIV-2 PR as presented here was completed before any such structure or crystallographic data had been published.

MATERIALS AND METHODS.

Cloning and Expression.

The HIV-2_{ROD} strain (Cheng, Patterson *et al.* 1992) was used as the source material for this study. An infectious DNA clone of HIV-2_{ROD} (designated pROD35B) had previously been prepared (by Larder, B. A. and Darby, G. at Wellcome laboratories Beckenham), from which we excised a HindIII / EcoRI region comprising the PR gene and flanking sequence (fig. 1a). Using this as a template, and with specifically engineered primers, we generated the PR gene as a 300 bp PCR product incorporating an Nco1 site with an integral ATG codon in frame with, and immediately preceding the PR coding sequence, plus a HindIII site with an integral TAA stop codon immediately following the gene sequence (fig. 1b). The PCR product was digested with both Nco1 and HindIII enabling ligation (fig. 1c) into a pKU2 LacIq plasmid vector (described in Chapter 3), to produce pKU2_{LacIq}(HIV-2PR). The new construct was subsequently used to transform *E.coli* MC1061 cells for protein expression.

Clones containing the correct plasmid were identified by restriction digests (fig. 2), induced to express the protease by addition of IPTG, and analysed for protease expression by Western blots with human HIV-2 polyclonal antisera as the primary antibody (fig. 3). The plasmid DNA sequence from those clones reacting positively with the HIV-2 antisera was determined by the Sanger di-deoxy method (Sanger, Nicklen *et al.* 1977; Sanger and Coulson 1978), as described in Chapter 2 (pages 130-133).

Cell growth and expression of recombinant HIV-2 Protease.

Cell growth and HIV-2 PR expression were first optimised on a small scale, and it was established that a regime essentially identical to that used to express HIV-1 PR (Chapter 3) provided the optimum yield of active enzyme. Typically a cell pellet of 2.5 gL^{-1} was obtained, and routinely the cell paste was stored in 20 g batches, at -70°C, in 50 ml of Buffer A (Chapter 3, table1, page 186) containing 2 mM PMSF.

Purification of recombinant HIV-2 Protease

A 20 g pellet of cells was first thawed, then lysed by passing three times through a French Press at 4°C. The cell debris was removed from the lysate by centrifugation at an RCF of 30000 g for 1 hour at 4°C. The supernatent was further clarified by ultracentrifugation at an RCF of 141000 g for 1 hour at 4°C, then dialysed for 12 hours at 4°C against Buffer B (Chapter 3, table 1, page 186). A slight precipitate formed during dialysis and the solution was clarified by ultracentrifugation, as described above. The solution (50 ml) then underwent SEC, exactly as described for HIV-1 PR in Chapter 3, and 20 ml fractions were collected for analysis using the Scintillation Proximity Assay (SPA) (fig. 4) which was carried out as described in Chapter 2 (page 163).

Active fractions were pooled and dialysed for 8 hours at 4°C, against 5 L of 1 M guanidine in Buffer F (table 1), followed by dialysis against 5 L of 1 M Ammonium Sulphate in Buffer F for 12 hours at 4°C. A slight precipitate formed during these dialysis procedures, which was removed by ultracentrifugation as described above. The clarified supernatent (120 ml) was then loaded at a flow rate of 0.5 $mlmin^{-1}$ onto a Pharmacia HiLoad 26/10 phenyl-sepharose HIC column previously equilibrated with a 1 M solution of ammonium sulphate in Buffer F. Protein was eluted using a gradient of ammonium sulphate (from 1 to 0 M) in Buffer F over a total of 50 ml (fig. 5).

Protein fractions were collected as peaks, analysed by SDS-PAGE (fig. 6) and assayed for activity using the SPA (fig. 7).

Active fractions were pooled (~ 10 ml), dialysed against Buffer G (table 1) for 12 hours at 4°C, and then loaded at a rate of 1 $mlmin^{-1}$ onto a 16/10 fast flow S-Sepharose IEC column, previously equilibrated with Buffer G.

Table 1 Buffers F, G and H

Buffer F:
100 mM Phosphate buffer pH 7.4, 1 mM EDTA, 1 mM DTT
Buffer G:
50 mM Na-acetate pH 4.4, 1 mM EDTA, 1 mM DTT, 10% v/v glycerol
Buffer H:
10 mM acetate pH 4.4, 1 mM EDTA, 1 mM DTT

The protease was eluted using a gradient of 0–1 M NaCl in Buffer G, with a stepped interval in the salt gradient at 400 mM NaCl (fig. 8). Peak fractions were collected and checked for activity by SPA (fig. 9); the active fractions were pooled and analysed by SDS-PAGE (fig. 10). Protease eluted during the 400 mM NaCl step and was essentially pure.

For crystallisation purposes, the pooled active fraction was dialysed against Buffer H (table 1) for 12 hours at 4°C. The volume of the protease solution was then reduced using a Centricon 10, until a concentration of approximately 15 $mgml^{-1}$ was reached (as determined by absorbance measurement at λ = 280 nm, given an extinction coefficient calculated to be 19060 $cm^{-1}M^{-1}$ using the combined methods of Edelhoch 1967 and Gill, and von Hippel 1989). Protein was analysed by SDS-PAGE at this final stage of the procedure to assess its purity (fig. 10).

Isoelectric focusing

The pI of HIV-2 PR was determined using an Immobiline DryStrip Kit, as directed by the manufacturer (supplied by Pharmacia). A pH range 3.0 - 10.5 was used in conjunction with a High pI calibration Kit (pH 5- 10.5). The running time was 16 hours, giving a total 22 kVh. Detection was achieved using Coomassie staining.

Crystallisation and data collection.

Crystallographic methods have been described in Chapter 2 (pages 170-171). The initial conditions for crystallisation were determined using a 'wild screen' in droplets under oil, and optimised by means of the hanging drop vapour diffusion method. Optimally crystals were grown over 3 to 5 days at a constant temperature of 15°C from hanging drops prepared using 3 μl of HIV-2 PR stock at 6 $mgml^{-1}$, mixed with an equal volume of a precipitating solution Buffer I (table 2), 1 ml of which was also used to fill the reservoir. Crystals grew to a diameter of 100 μm at best, and smaller crystals were macroseeded into fresh drops to augment their size.

Table 2 Buffer I.

0.1 M NaHEPES pH 7.5, 0.2 M $CaCl_2$, 10% v/v polyethylene glycol (PEG) 400

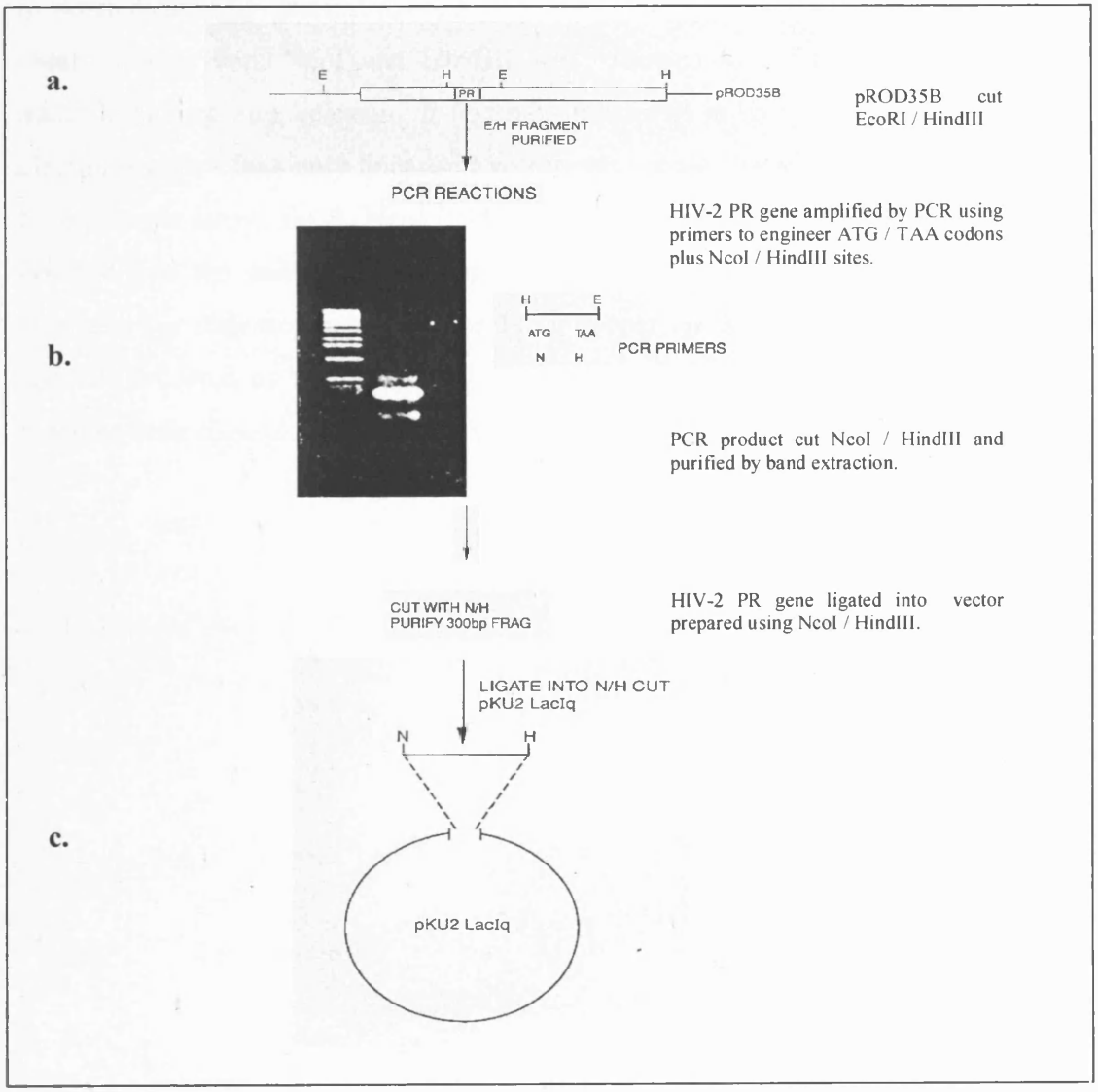
Crystals were mounted into glass capillaries as described in Chapter 2 (Crystallographic Methods, page 174). Crystals were analysed by X-rays at the Daresbury SRS, the data collection and processing procedures are also described in Chapter 2 (pages 175-178).

RESULTS AND DISCUSSION.

Cloning and sequencing of the HIV-2 PR gene.

The HIV-2 PR gene was isolated by PCR from an EcoRI / HindIII DNA fragment derived from the construct pROD35B (fig. 1). NheI and HindIII restriction enzyme sites were engineered into the 5′ and 3′ termini of the gene during the PCR step allowing insertion into the expression vector $pKU2_{LacIq}$ (as used previously with HIV-1 PR - see Chapter 3).

Figure 1 Construction of $pKU2_{LacIq}$ HIV-2PR.



The new construct was used to transform *E.coli* MC1061 and miniprep DNA was isolated from these transformants for analysis. Two clones, numbered 11 and 12, were identified by restriction analyses using the enzymes NcoI, HindIII and SspI (fig. 2). The vector alone contained two SspI sites and thus digestion of the control released an SspI-SspI fragment in addition to the linearised vector (indicated by the blue arrow in fig. 2, SSP 1, lane C). Since the PR gene also contained an SspI site any construct containing the PR gene should release two SspI-SspI fragments in addition to the linearised vector on digestion with this enzyme – as clones 11 and 12 were observed to do (indicated by green arrows, fig. 2, SSP 1, lanes 11 and 12). Digestion of the control vector with NcoI and HindIII was expected to cut the vector within its multiple cloning site, releasing a fragment too small to be resolved by agarose gel electrophoresis - thus only linearised vector was visible in the control lane (indicated by the purple arrow, fig. 2, NcoI / HindIII, lane C). However, since the PR gene was inserted into the construct via NcoI and HindIII sites, digestion with both these enzymes was expected to release the PR fragment, as was observed with clones 11 and 12 (indicated by the red arrow, fig. 2, NcoI / HindIII, lanes 11 and 12). Taken together these digests verified the presence of the PR gene fragment in clones 11 and 12.

Figure 2 Restriction enzyme analysis of pKU2$_{LacIq}$ (HIV-2 PR) construct.

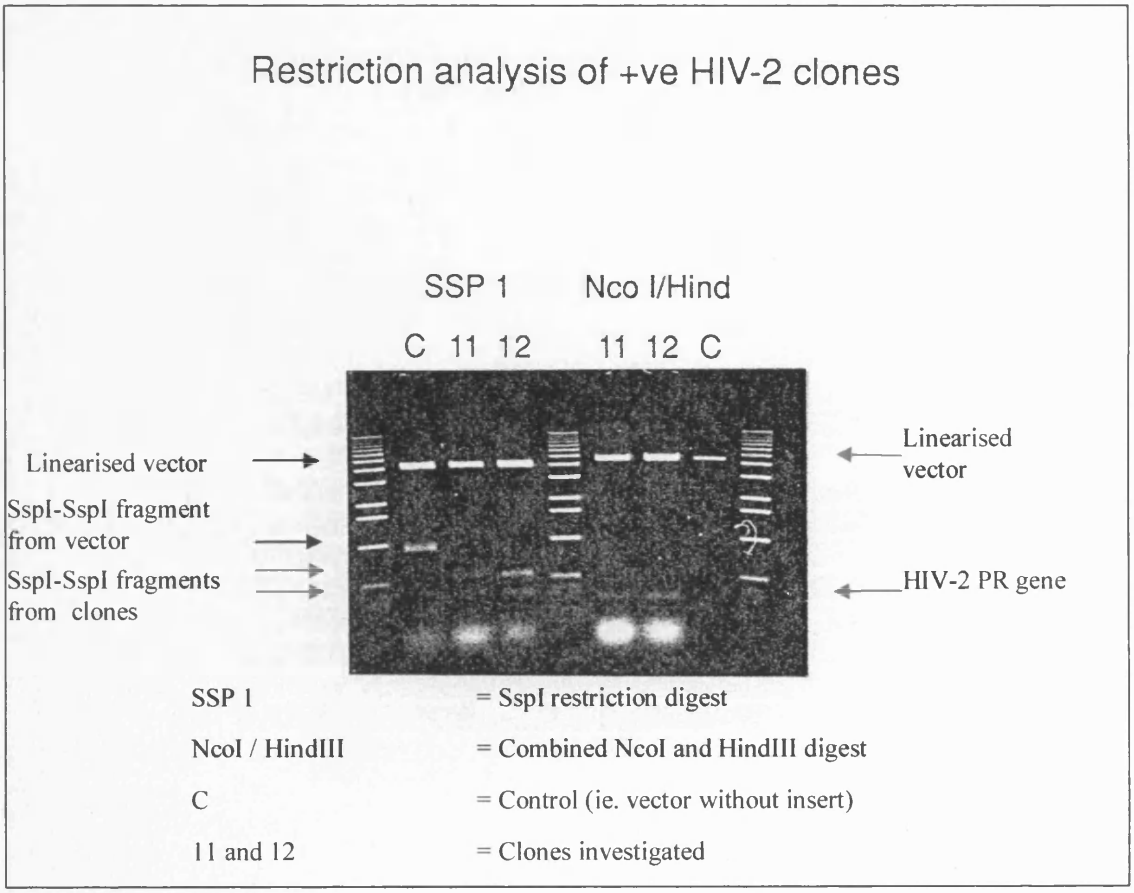
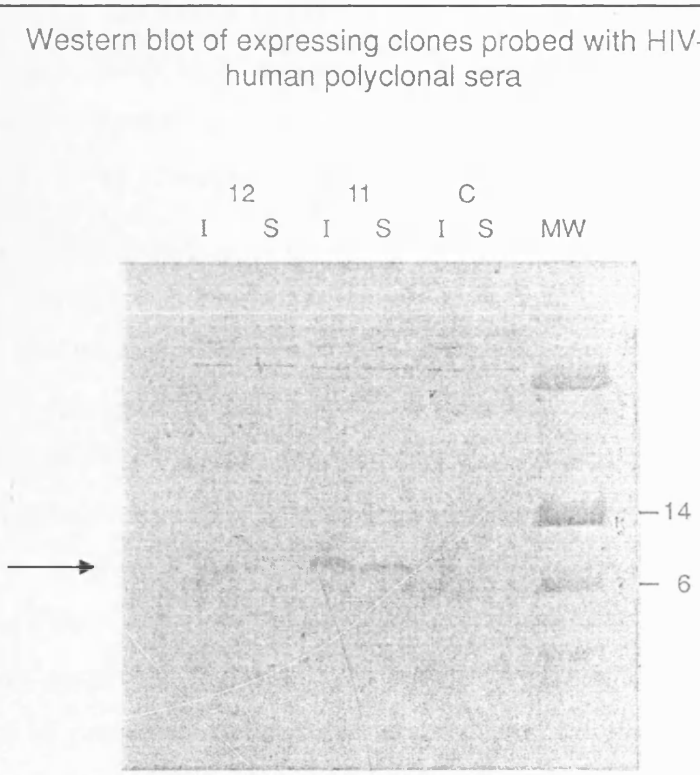


Figure 3 Expressed protein cross reacts with HIV-2 PR anti-sera.



C	Control (*E.coli* lysate in the absence of HIV-2 PR construct)	
MW	Molecular weight markers (indicated in kD)	
I	Insoluble fractions	11 = lanes relating to clone 11
S	Soluble fractions	12 = lanes relating to clone 12

The arrow indicates the position of the HIV-2 PR bands. These are most clearly seen in lanes relating to clone 11, and particularly in the insoluble fraction derived from clone 11. No equivalent band is seen in the control lanes.

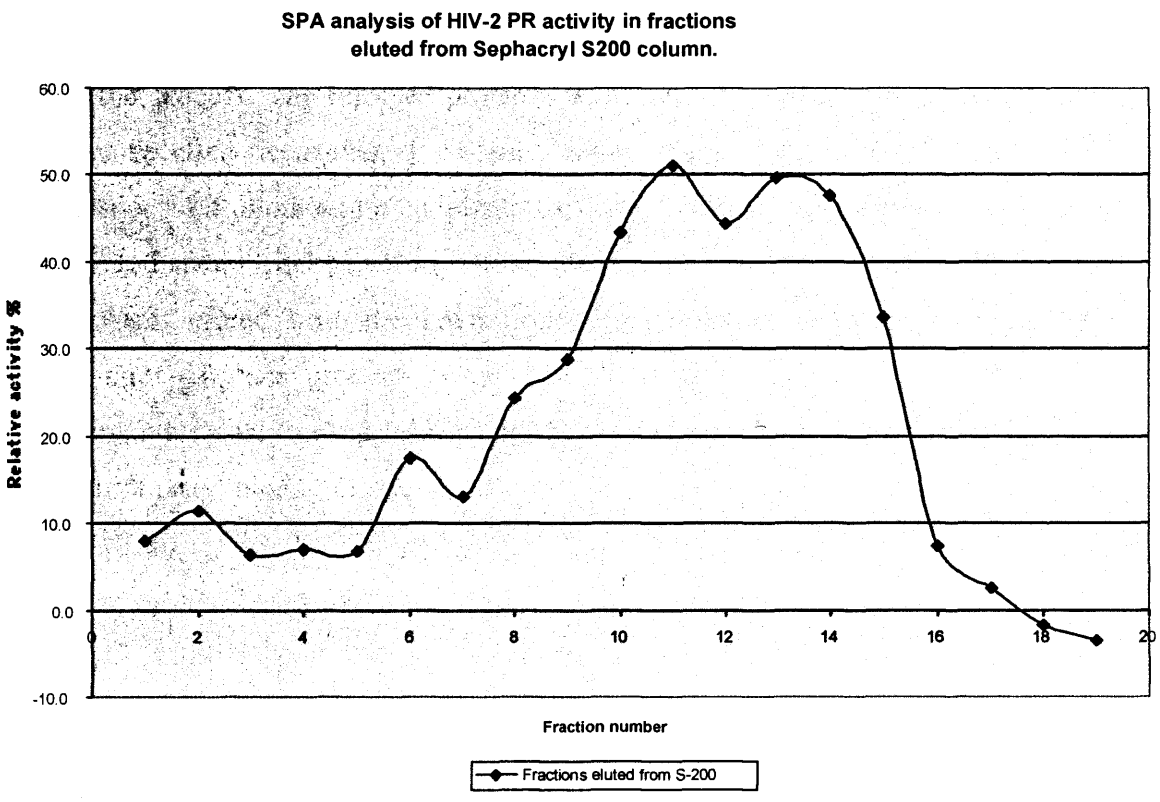
Western blot analysis of proteins expressed after IPTG induction of the positive clones showed that both clones expressed a soluble protein of molecular weight ~ 10 kD, which cross reacted with HIV-2 human polyclonal anti-sera (fig. 3).

DNA sequencing of both clones 11 and 12 showed 4 base substitutions, when compared with the pROD35B sequence, although none of these substitutions would have resulted in a change to the amino acid sequence. In addition, clone 12 had further base changes which resulted in the alteration of one amino acid, consequently clone 11 was selected for all subsequent work.

Purification of HIV-2 PR.

Although HIV-2 PR was produced solubly in *E.coli*, it was necessary to employ denaturing conditions for the initial SEC, since it was found that under native conditions the protein did not resolve on this column. It had previously been established with HIV-1 PR that under denaturing conditions it was possible to resolve the protein from the Sephacryl S-200 column, resulting in a 60% purification of the protease. Furthermore it had then been possible to refold it into an active, crystallisable form (Chapter 3). It was anticipated that HIV-2 PR could be prepared in the same way. Following dialysis of the soluble extract of HIV-2 PR into Buffer B (containing 5 M guanidine) and chromatography using a Sephacryl S-200 column, the protease eluted along with the low molecular weight peak, and active fractions were detected by means of SPA (fig. 4). The SPA data is presented in table 3 of the appendix (page 345).

Figure 4 HIV-2 PR activity detected in Sephacryl S200 elution fractions.



Fraction volumes = 20 ml

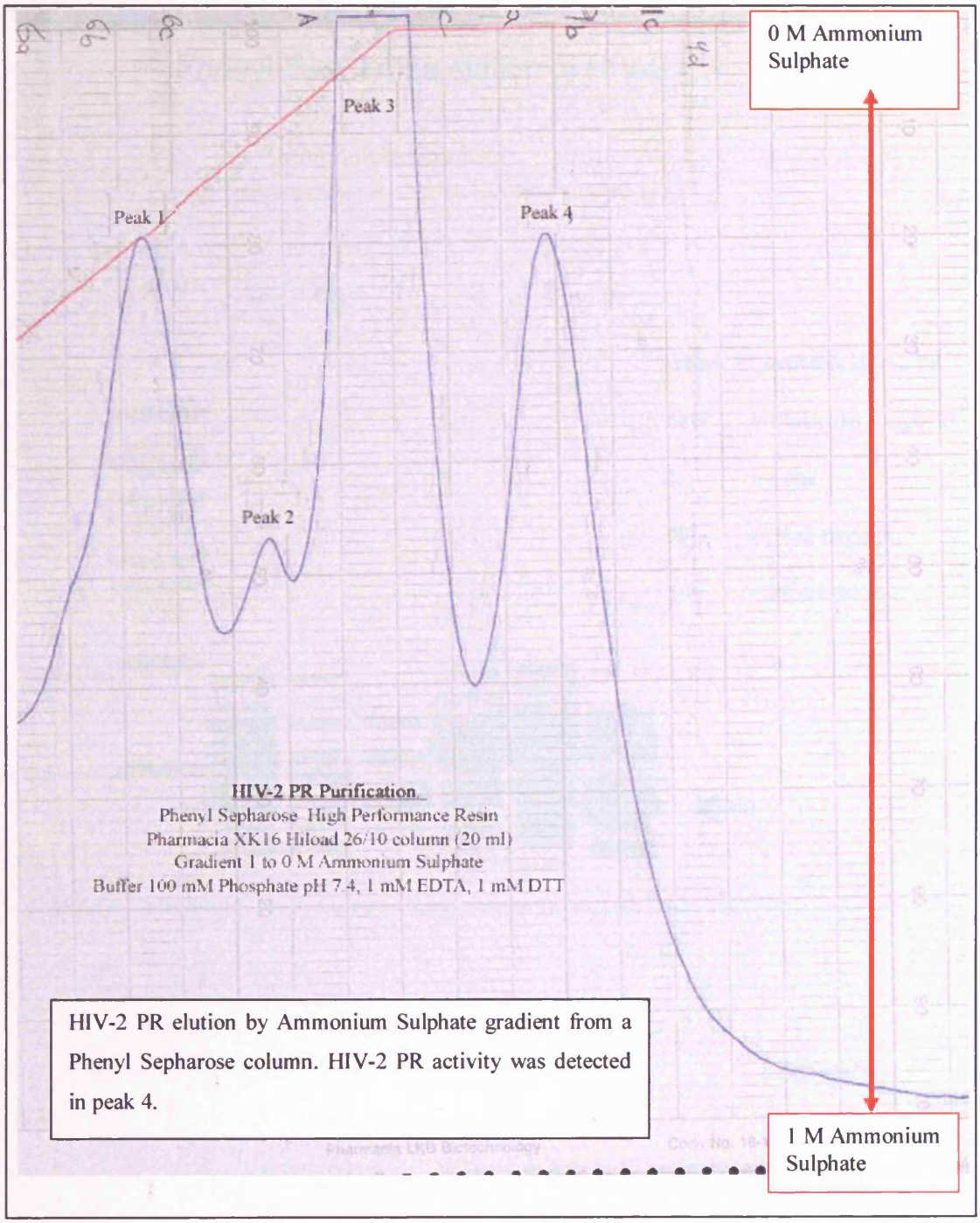
Fractions 10-15 contained HIV-2 PR activity and were pooled.

Column = XK50, column volume = 2000 ml.

Dialysis of the eluent against 1 M guanidine in Buffer F (table 1) resulted in the refolding and further purification of the protease. It also served to replace an MES buffer (Buffer B – Chapter 3, table 1) with a phosphate buffer (Buffer F), raising the pH in the process. The refolded protease was then prepared for the phenyl-sepharose column by dialysis against 1 M ammonium sulphate in Buffer F. By replacing guanidine with ammonium sulphate this step caused the protease to adopt its dehydrated state - enabling it to bind to the phenyl-sepharose column HIC column[11]. Protein was eluted from the Phenyl-Sepharose column by means of a decreasing ammonium sulphate gradient, from 1 to 0 M ammonium sulphate, (fig. 5) and fractions were analysed by PAGE (fig. 6). Protease activity was detected by SPA in the fourth peak, having eluted at 0 M ammonium sulphate (fig. 7 and appendix table 4, page 346).

[11] The two-state protein model (Chen and Sun 2003) assumes that a protein can exist in two states - hydrated and dehydrated. Normally a protein is coated with structured water molecules, forming a hydration shell over the protein surface (the hydrated state). The addition of salt (usually ammonium sulphate) to a protein solution strips the hydrophobic surfaces of a protein of its hydration shell, causing the protein to adopt the dehydrated state in which the hydrophobic zones of its surface are exposed to the bulk solvent. These hydrophobic patches then tend to associate with other hydrophobic ligands. This association is termed "hydrophobic interaction", and forms the basis for this separation technique.

Figure 5 Phenyl-Sepharose Chromatograph.



HIV-2 PR elution by Ammonium Sulphate gradient from a Phenyl Sepharose column. HIV-2 PR activity was detected in peak 4.

Figure 6 Analysis of Phenyl-Sepharose fractions by PAGE.

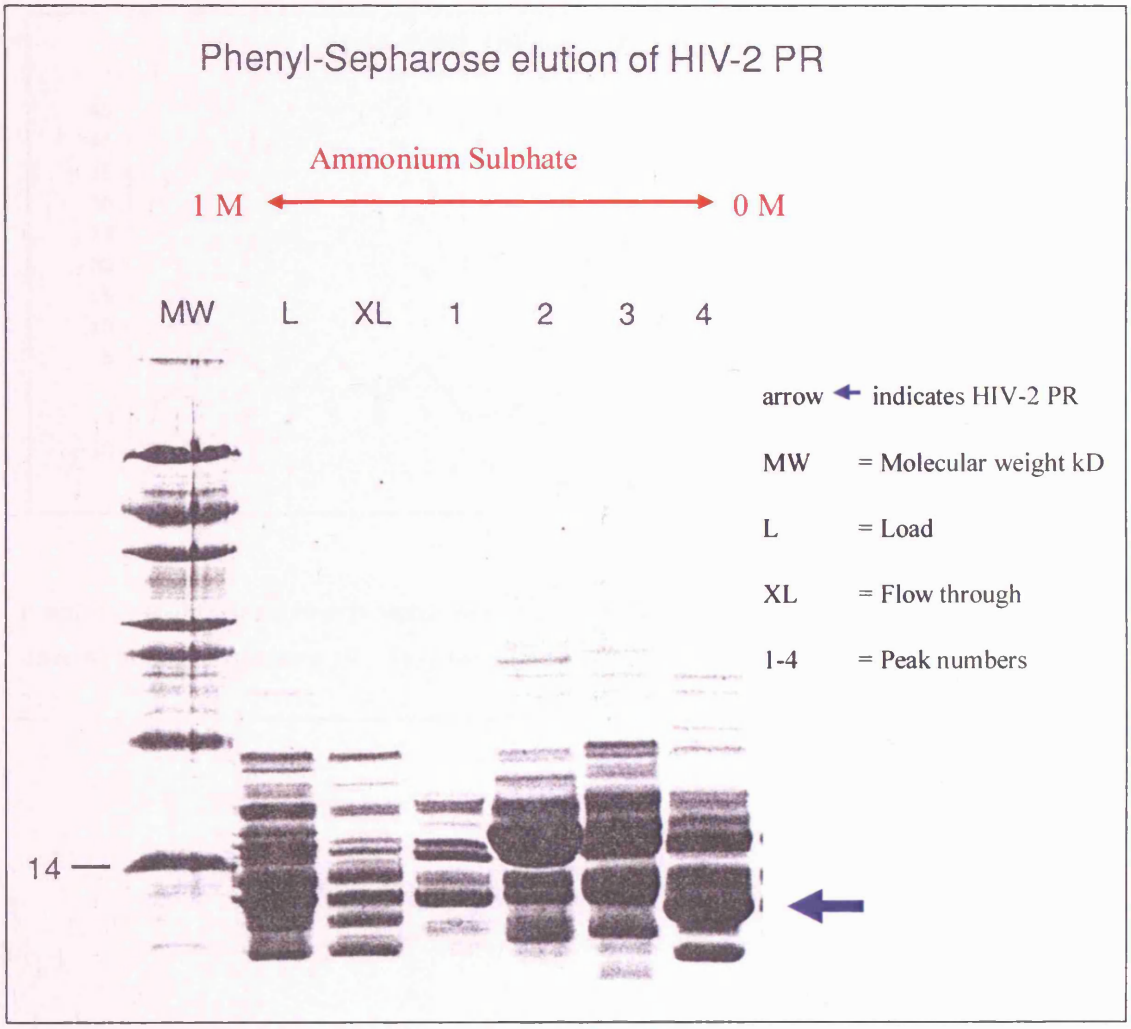
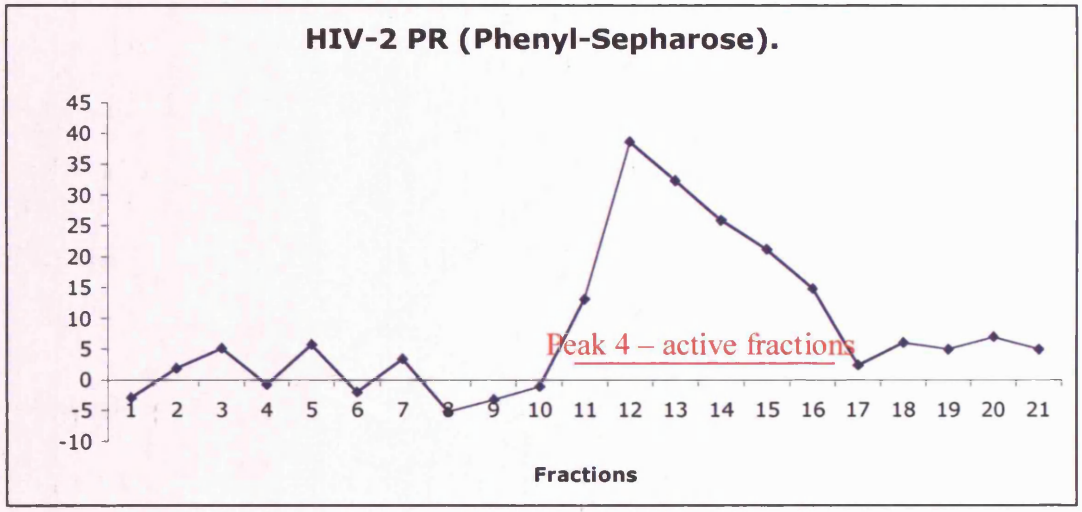
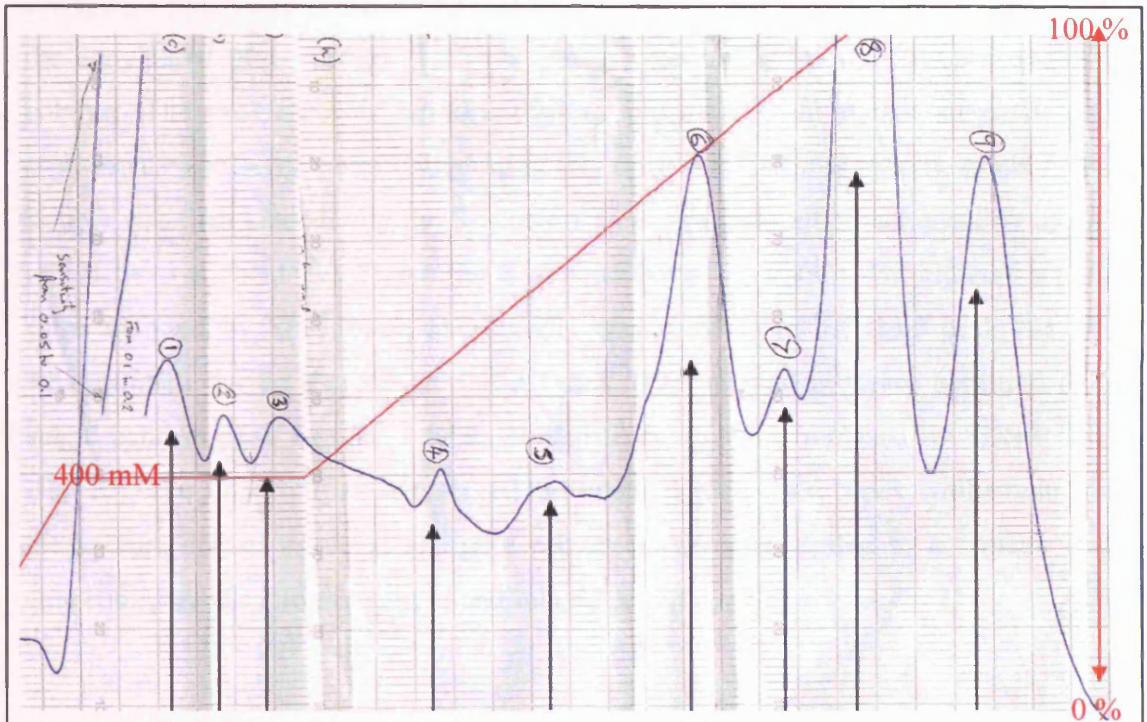


Figure 7 Detection of active fractions eluted from Phenyl Sepharose



Fractions eluted from the Phenyl Sepharose column were subjected to SPA and HIV-2 PR activity was detected in peak 4 (fractions 10 – 16) - see appendix table 4, page 346; and figs. 5 & 6, pages 216 & 217.

Figure 8 Chromatograph of HIV-2 PR elution from S-Sepharose.



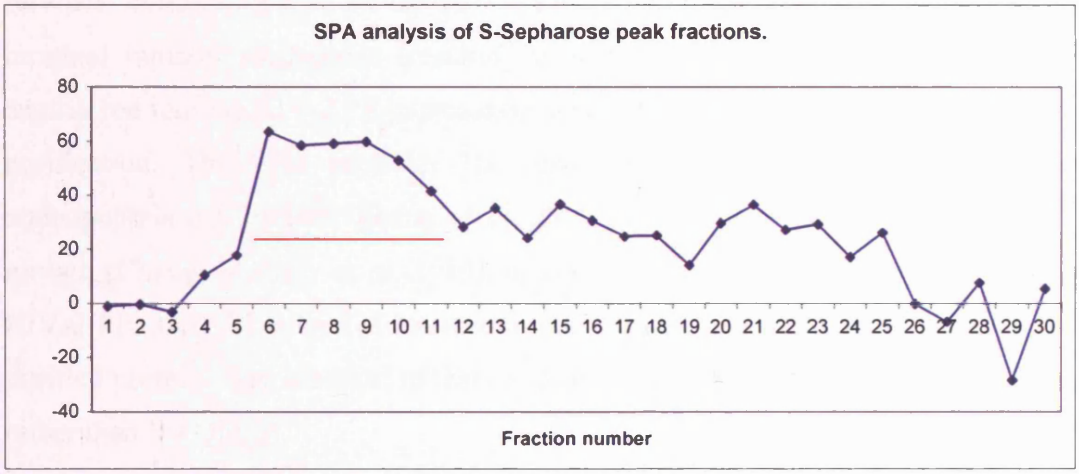
= UV Absorbance (280 nm)

= % NaCl gradient (100 % = 1 M NaCl) with step at 400 mM NaCl.

HIV-2 PR was detected in peak 3.

Peak four was then dialysed against Buffer G, lowering the pH from 7.4 to 4.4 by changing a phosphate buffer for an acetate buffer. This ensured that the pH was below the anticipated pI (~ pH 5) of HIV-2 PR in preparation for the S-Sepharose IEC column. The protease was not observed to precipitate out at this stage, despite crossing its isoelectric point - suggesting that once refolded it is readily soluble. The sample was then loaded onto the S-Sepharose column and elution achieved by means of a stepped gradient (fig. 8) in order to obtain a better separation. The eluted fractions were analysed by SPA assay, and activity was detected in the third peak (fig. 9), which eluted during the 400 mM NaCl step. The purified protease appeared as a single band in peak number 3 when pooled fractions were analysed by SDS-PAGE (fig. 10). The final dialysis step served to remove glycerol and NaCl, whilst reducing the sodium acetate content. The pure protease was concentrated to a typical final concentration of approximately 15 mg/ml and stored pending use at -70°C.

Figure 9 SPA analysis of S-Sepharose peak fractions.



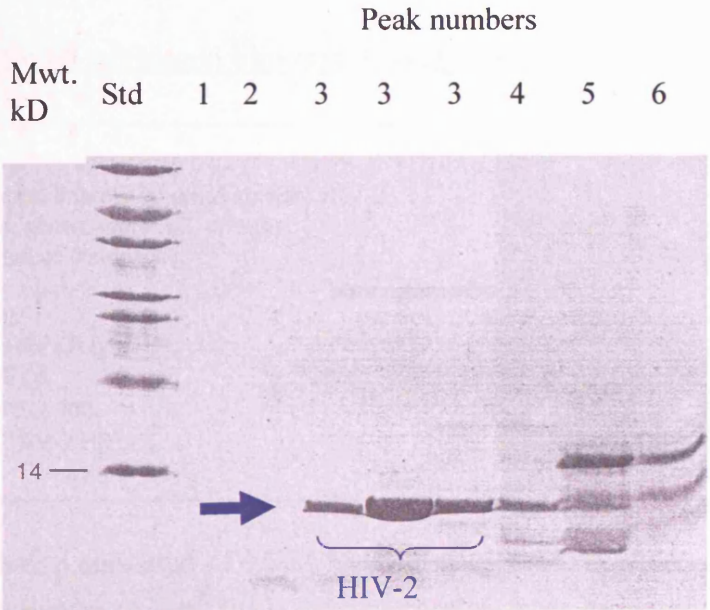
Peak number 3 (see fig. 8, page 219) incorporated fractions 6 – 15 inclusive.

Fractions 6 – 11 contained the bulk of PR activity and were pooled

Data shown in appendix table 5, page 347.

Figure 10 SDS PAGE analysis of S-Sepharose peak fractions.

Peak numbers refer to peaks from S-Sepharose chromatograph shown in figure 8, page 219.



Characterisation of purified HIV-2 PR.

The purified protease was subjected to amino-terminal analysis of the first 19 residues, confirming it to be the HIV-2 PR desired. This also revealed that the N-terminal initiator methionine (resulting from translation of the ATG start codon, engineered into the HIV-2 PR expresssion vector) had been lost during expression and purification. This was probably the result of cleavage by *E.coli* methionine aminopeptidase (Lowther, Zhang *et al.* 1999), an enzyme essential for *E.coli* cell growth (Chang, McGary *et al.* 1989), or conceivably as a result of cleavage by the HIV-2 PR itself. This loss of the methionine ensured that the amino-terminal of the purified protease was identical to that of naturally occuring HIV-2 PR (ie. PQFSL etc rather than MPQFSL).

Isoelectric focusing revealed that the pure active HIV-2 PR has a pI of 5.2.

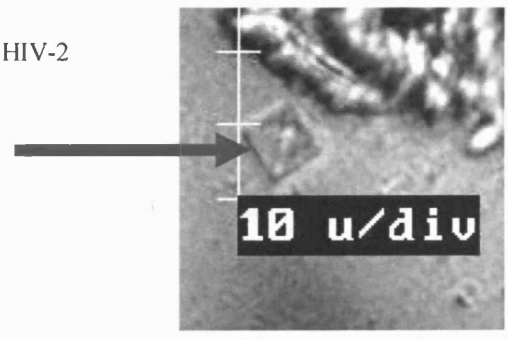
Crystallisation, data collection and analysis.

The crystal screen resulted in the growth of small tetragonal bipyramids under two separate, though similar, conditions. The first of these consisted of 0.1 M NaHEPES buffer (pH7.5) containing 0.2 M $CaCl_2$, and 20% v/v PEG 400. This solution resulted in greater than 20 tetragonal bipyramids from a 2 µl drop under oil - at least ten of which were single, the others being twinned. The first of the bipyramids appeared after a period of 5 days, the remainder developed over a period of one month.

Figure 11 Small tetragonal bipyramid crystals were observed under oil.

Small (approximately 10 µm diameter) HIV-2 PR crystal, grown under oil, prior to optimisation of conditions.

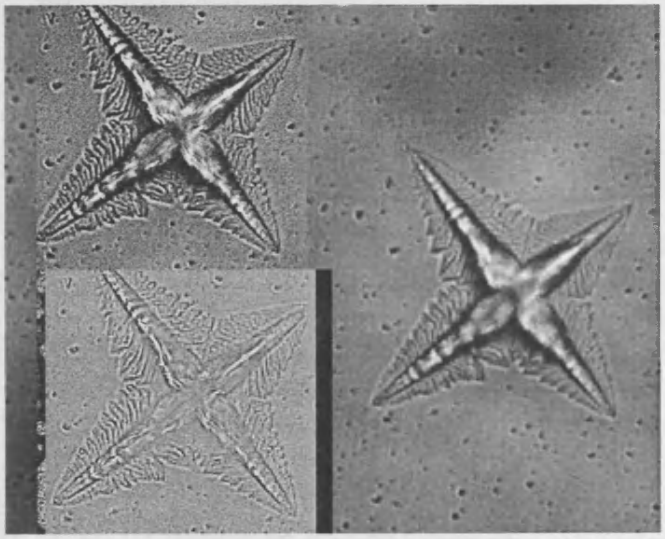
Conditions:
0.1 M NaHEPES (pH7.5)
0.2 M $CaCl_2$
20% v/v PEG 400
1 mg/ml [HIV-2 PR]



The second solution consisted of 0.1 M NaHEPES (pH 7.5) containing 0.2 M $MgCl_2$ and 30% v/v PEG 400. This solution produced clusters of very small tetragonal bipyramids (identical in morphology to those grown from the first condition) after a period of one month.

Other solutions screened which incorporated either a NaHEPES buffer (pH 7.5), PEG 400, a group two metal salt, or a chloride were reviewed, though no further crystals were observed. A series of hanging drop experiments were carried out to investigate optimum conditions for crystal growth using both solutions. Initial investigation centred on protein concentration. A final protein concentration of > 3 $mgml^{-1}$ tended to produce very rapid crystal growth, and crystals formed were of poor morphology - often feathered or twinned (fig. 12) - therefore 3 $mgml^{-1}$ was adopted as the maximum final concentration for further study.

Figure 12 Poor crystal morphology was observed at [HIV-2 PR] > 3 mg/ml.

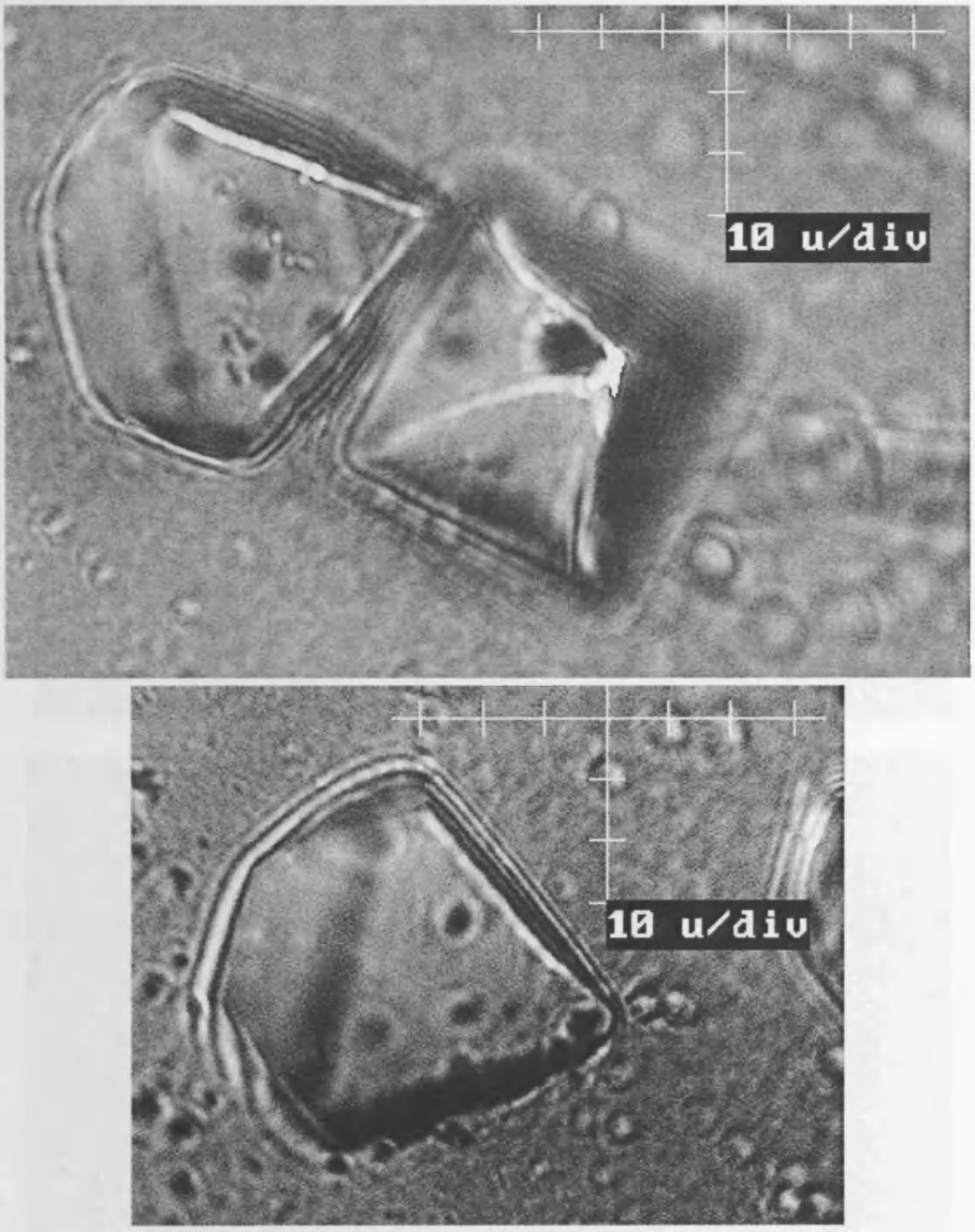


Example of poor crystal morphology.

Conditions:
0.1 M NaHEPES (pH 7.5)
0.2 M $MgCl_2$
30% v/v PEG 400
5 mg/ml HIV-2 PR
Vapour diffusion at 20°C for 14 hours.

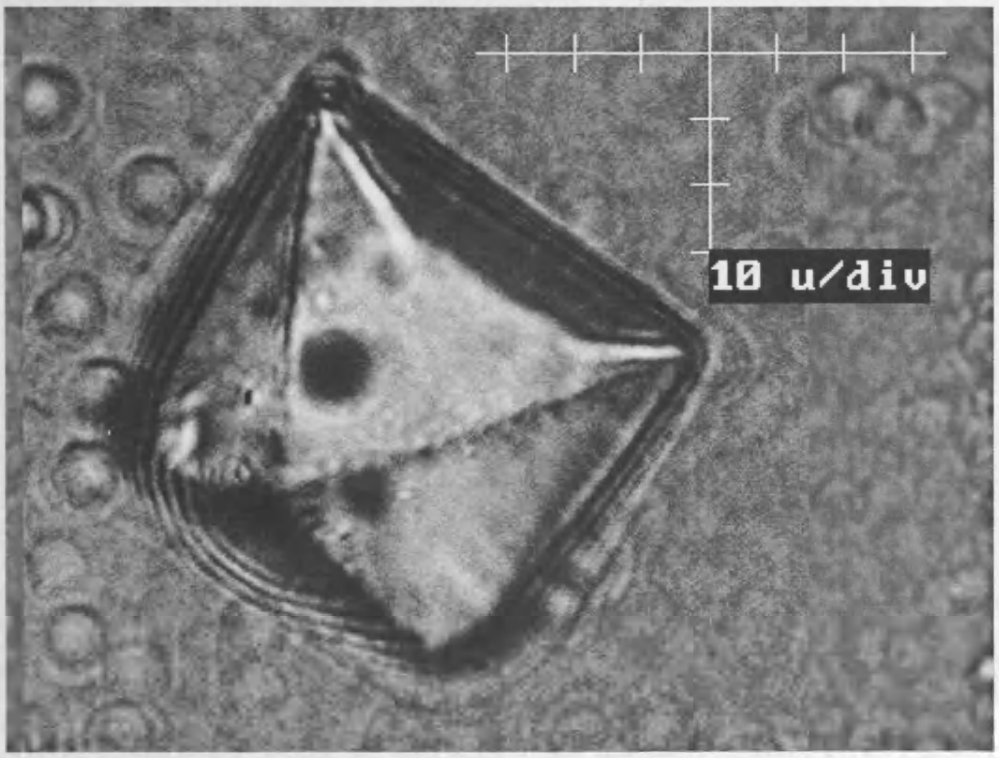
The effect of a lowered PEG 400 concentration was investigated, as was the effect of the extra ionic components. A series of hanging drops containing either $CaCl_2$, $MgCl_2$, NaCl, or no ionic component, and using varying concentrations of PEG 400 over the range 5% v/v to 10% v/v was set up. Following 52 hours at 20°C 12 single tetragonal bipyramids of between 50 - 80 μm in length (figs. 13 and 14) were observed in the drop containing a final PEG 400 concentration of 10% v/v, and $CaCl_2$ at 0.2 M. No crystals were observed under the other conditions. Repetition of these conditions yielded further tetragonal bipyramids of 100 μM in length (fig. 15).

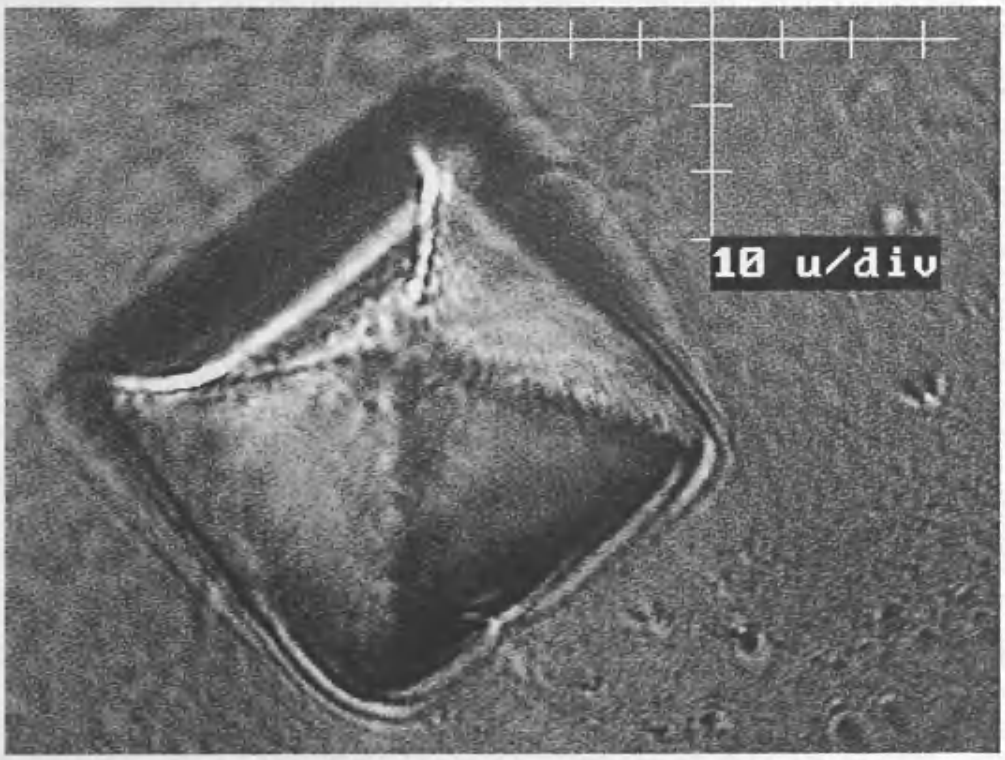
Figure 13 Tetragonal crystals of HIV-2 PR grown by vapour diffusion.



Crystals shown above are approximately 50 μm in diameter and were grown by vapour diffusion from 0.1 M NaHEPES (pH 7.5), 10% v/v PEG 400, and 0.2 M $CaCl_2$ at 20°C over a period of 52 hours.

Figure 14 Optimisation of HIV-2 PR crystals grown by vapour diffusion.

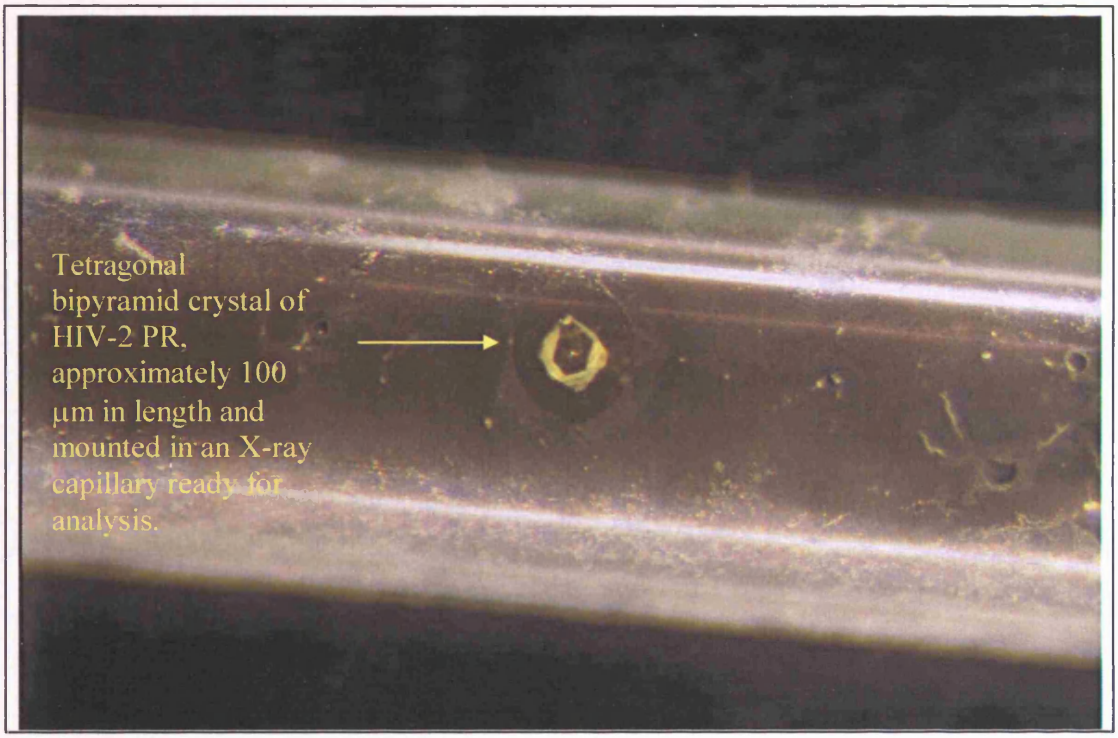




Crystals of approximately 80 μm diameter were grown by vapour diffusion from:

0.1 M NaHEPES (pH 7.5), 10% v/v PEG 400, and 0.2 M $CaCl_2$ at 20°C over 52 hours.

Figure 15 HIV-2 PR crystal mounted for X-ray ananlysis.

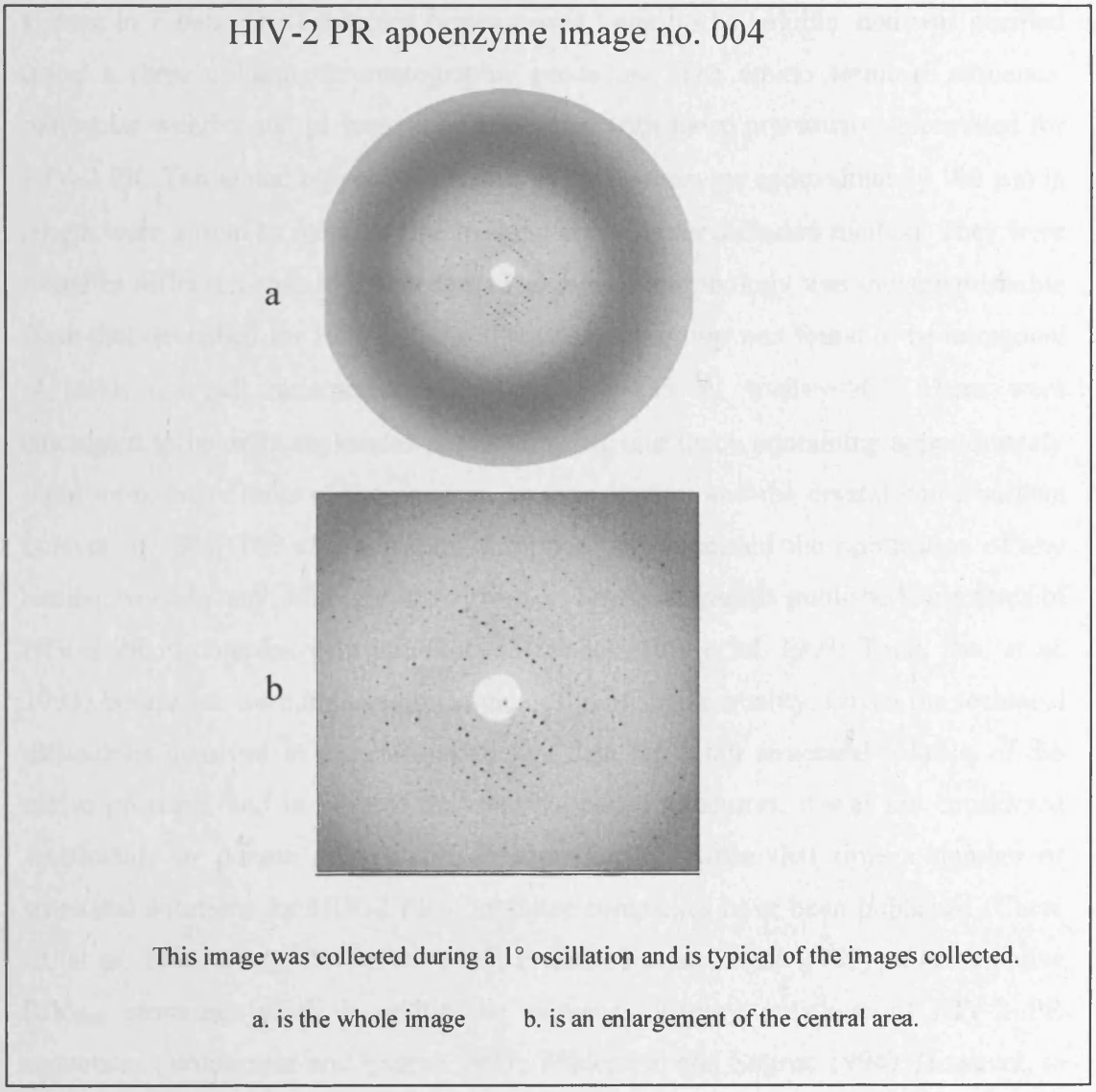


The pH of crystallisation buffer established here reflects the results of McKeever *et al.*, who found that well formed tetragonal bipyramids of HIV-1 PR were obtained at pH 7 only. The morphology of the HIV-2 PR crystals described here was essentially indistinguishable from that of HIV-1 PR crystals (see Chapter 3, fig.9, page197). The crystals were analysed by SDS-PAGE and Western blot, and the enzymatic activity of re-dissolved crystal protein was investigated using the SPA method. In both cases results were identical to those obtained with the stock protein (data not shown).

The crystals were found to diffract to between 3 and 4 Å using the SRS at Daresbury (December 1992), although they were not stable long enough for significant data collection. This was probably a function of their small size and possibly indicative of high water content. Attempts were made to increase crystal size by macro seeding techniques (described Chapter 2, page 172) which eventually yielded crystals with diameters > 100 μm. However, no significant improvement in stability was seen, despite the increased crystal size. Stability might have been improved by co-crystallisation with an inhibitor. This was attempted using pepstatin, but without success - no crystals were observed in the presence of pepstatin.

A number of macroseeded crystals were mounted for x-ray analysis (fig. 11), and exposed to x-rays both at Birkbeck's in house source, and at the Daresbury SRS. The best results were obtained at the SRS (in April of 1994) and consisted of only three stills (at 0°, 90° and 45°), and a further 9 oscillations (fig. 12) of 1° per oscillation (starting at 0°) before data collection was abandonned due to deterioration of the crystal. The crystal diffracted to 4.5 angstroms resolution, and was found to be in a tetragonal space group I4, with unit-cell parameters of a=b=130, c=185 Å, $\alpha=\beta=\gamma=90°$. There were calculated to be eight molecules per asymmetric unit (each containing approximately eight monomeric units of the enzyme, or four dimers) and the crystal had a solvent content of 73%. Typically the solvent content of a crystal may range from between 25% to 70 %, meaning that the water content of these crystals was very high. This coupled with the complexity of the asymmetric unit, and the extreme difficulty involved in collecting sufficient data (at nine oscillations per crystal, forty crystals would be required to complete an entire data set) it was considered that to pursue a structural solution was unfeasible without the use of specialised techniques (such as Cryo-crystallography and solvent flattening). Cryo-crystallography in particular may have improved chances of obtaining sufficient data since this technique minimises damage to the crystal during data collection by freezing it throughout the process in a stream of Nitrogen. Consequently individual crystals diffract for longer periods, and fewer are therefore required to complete a data set (often one is sufficient). However, this technique was in its infancy at that time, and was not available to us. The only viable alternative was to obtain crystals of HIV-2 in complex with an inhibitor, in the hope that the presence of the inhibitor would stabilise the crystal and allow it to diffract more efficiently and for longer periods. However, in light of the two contemporary structural solutions for HIV-2 – inhibitor complexes that had recently been published (Mulichak, Hui, *et al.* 1993; Tong, Pav, *et al.* 1993) it did not seem worthwhile to pursue this approach.

Figure 16 HIV-2 PR diffraction pattern.



This image was collected during a 1° oscillation and is typical of the images collected.

a. is the whole image b. is an enlargement of the central area.

SUMMARY

The HIV-2 PR was cloned, expressed and purified from a recombinant expression system in *E.coli.* The expressed protease was found to be soluble, and was purified using a three column chromatographic procedure. The amino terminal sequence, molecular weight, and pI were shown to agree with those previously determined for HIV-2 PR. Tetragonal bipyramid crystals of the apoenzyme approximately 100 μm in length were grown by means of the hanging drop vapour diffusion method. They were found to diffract x-rays to 3 Å at best. The crystal morphology was indistinguishable from that described for HIV-1 PR, whilst the space group was found to be tetragonal I4, with unit-cell parameters of a=b=130, c=185 Å, $\alpha=\beta=\gamma=90°$. There were calculated to be eight molecules per asymmetric unit (each containing approximately eight monomeric units of the enzyme, or four dimers) and the crystal had a solvent content of 73%. The crystallisation described here preceded the publication of any similar work by any other groups. However, two rival groups published structures of HIV-2 PR in complex with inhibitors (Mulichak, Hui, *et al.* 1993; Tong, Pav, *et al.* 1993) before we were able to obtain data of publishable quality. Given the technical difficulties involved in obtaining sufficient data for a full structural solution of the native protease, and in light of the rival published structures, it was not considered worthwhile to pursue a structural solution further. Since that time a number of structural solutions for HIV-2 PR – inhibitor complexes have been published (Chen, Li, *et al.* 1994, Tong, Pav, *et al.* 1995, Priestle, Fassler, *et al.* 1995) as has a native SIV_{Mac} structure which is within the sequence identity envelope of HIV-2 PR sequences (Wilderspin and Sugrue 1993; Wilderspin and Sugrue 1994). However, to date no structure for the native HIV-2 PR has been published or deposited. The similarity of SIV_{Mac} and HIV-2 proteases, and the availability of the native SIV_{Mac} sequence for modelling purposes have arguably made a native HIV-2 PR structure less interesting, and therefore less likely to be pursued in the future. The native SIV_{AGM} PR structure by contrast has possibly become more interesting, given the relative dissimilarity of SIV_{AGM} PR to either HIV-1 or HIV-2 proteases (see Introduction, fig. 2, page 27) demonstrating it to be a truly Simian in origin.

While considering the PR structure and mechanism a general comparison of cysteine and aspartic protease active sites was made, leading to the proposal that their catalytic residues might be interchangeable. This hypothesis was investigated next (Chapter 5).

Chapter Five: The mutagenesis of HIV-2 Protease.

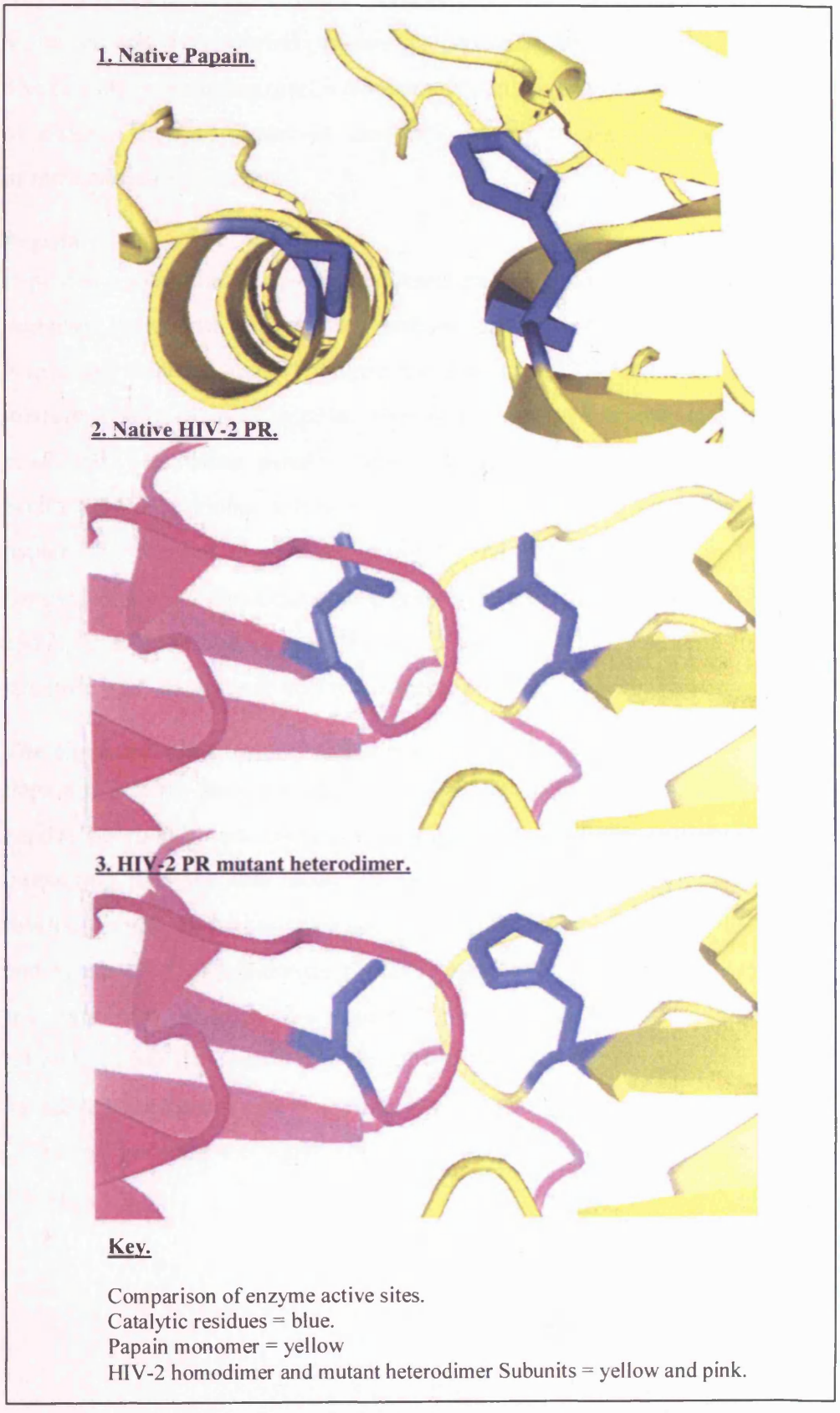
Three dimensional modelling, construction, expression, and purification of HIV-2 PR active site mutants, and the subsequent creation and characterisation of an HIV-2 PR heterodimer with the catalytic residues of a cysteine proteinase.

INTRODUCTION

A comparison of HIV-2 PR and the cysteine proteinase papain in terms of substrate specificity, mechanism of action, amino acid sequences and overall structures would reveal that they are very different. However, it was proposed that if the two catalytic aspartate residues of HIV-2 PR were replaced by the catalytic residues of papain (i.e. His and Cys) their relative spatial orientations in the resultant mutant HIV-2 PR protein would be sufficiently close to their native arrangement in the papain active site that the mutant HIV-2 PR might retain proteolytic activity (fig. 1). Catalytic activity would thus persist in the mutant enzyme despite replacement of the aspartates ordinarily essential for catalysis, because the replacement residues would enable proteolysis by means of a different mechanism - that of a cysteine proteinase rather than an aspartic proteinase. This chapter describes the modeling and creation of a heterodimeric HIV-2 PR His-Cys mutant in order to test this hypothesis[12] .

[12] In this Chapter the catalytic residues of the HIV-2 PR dimer are given as single letter / residue no. code and are separated by an oblique / , ie. the native homodimeric HIV-2 PR is D25/D25, the homodimeric HIV-2 PR mutants are C25/C25 and H25/H25 respectively, and the heterodimeric mutant is C25/H25. Numbering is as for HIV-PR.

Figure 1 Comparison of Papain and proposed mutant of HIV-2 PR



HIV-2 PR

HIV-2 PR is a dimeric, retroviral, aspartic protease. The characteristics of the aspartic proteases, and the retroviral proteases in particular, have been discussed in Chapter 1. The cloning, expression, purification and crystallisation of HIV-2 PR were the subject of chapter 4. The reader is therefore refered to these chapters as sources of introduction to this enyme.

Papain

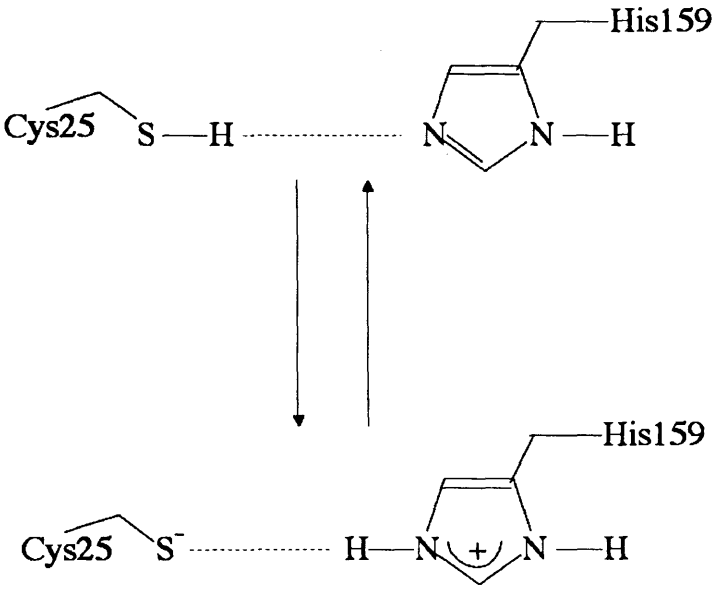
Papain is a monomeric, cysteine protease present in the tropical papaya fruit (*Carica papaya*). Proteolytic activity was first associated with extracts from the papaya by Wurtz and Bouchut in 1879 (Wurtz and Bouchat 1879), and was later attributed to a mixture of four or more cysteine proteases and other enzymes (Brocklehurst, Baines *et al.* 1981) including papain. Papain has been isolated and purified (Kimmel and Smith 1954; Brocklehurst, Baines *et al.* 1981) and extensively studied. Early work on papain is discussed in reviews by Smith and Kimmel (Smith and Kimmel 1960), Brocklehurst *et al.* (Brocklehurst, Baines *et al.* 1981; Brocklehurst, Willenbrock *et al.* 1987), and Baker and Drenth (Baker and Drenth 1987). Papain's three dimensional structure was the first solved for a cysteine protease (Drenth, Jansonius *et al.* 1968).

The biochemical characteristics of papain

Papain is a fairly basic protein with a pI of 8.75. Structurally it is a monomeric 212 residue polypeptide of 23429 Da. It is globular, has three disulphide bonds and two interacting domains that create the substrate-binding cleft. An active site cysteine residue (Cys25, Papain numbering) has been shown to be essential for activity (Finkle and Smith 1958). Furthermore Cys25 is required to be in a free state, and the active site must also incorporate a nearby histidine residue (His159, papain numbering) (Drenth, Jansonius *et al.* 1971; Glazer and Smith 1971; Drenth, Kalk *et al.* 1976). In its active state the enzyme is thought to consist of a thiolate-imidazolium ion pair (fig. 2) formed between the active site residues Cys 25 and His 159 (Storer and Menard 1994).

Figure 2 The active site catalytic residues of Papain.

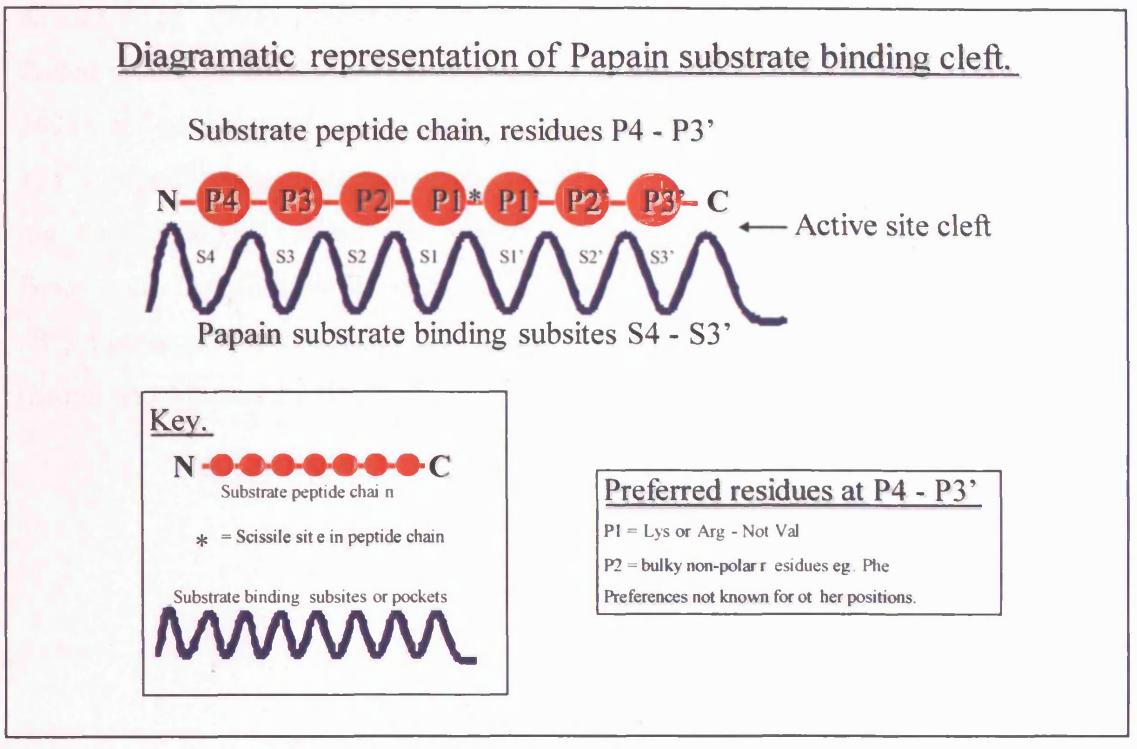
Thiolate - imidazolium ion pair between residues Cys25 & His159.



The catalytic Cys25 is involved in a tautomeric equilibrium between the neutral and zwitterionic forms.

Papain has seven subsites in its active site cleft (fig. 3), 4 (S1-S4) accomodate residues P1-P4 on the N-terminal side of the scissile site and 3 (S1'-S3') accept residues that are C terminal of it (P1'-P3'). Arg and Lys are prefered at position P1 (Kimmel and Smith 1957) while Val is not accepted (Baker and Drenth 1987). Bulky non-polar side chains are prefered at P2 - eg Phe (Berger and Schechter 1970). Little is known about the preferences of the S' subsites, but the choice of residues at P1'-P3' appears to be broad (Menard, Carmona *et al.* 1993). The lack of any clear preferences in terms of substrate binding means that papain can generate numerous cleavage patterns, and papain sites within a protein are hard to predict.

Figure 3 The papain substrate binding cleft.



The enzyme is stable and active between pH 4 and pH 10, and at temperatures up to 80°C (Glazer and Smith 1971), and still retains activity in the presence of 8 M urea. It can be stored at 4°C, but will lose up to 50% of its activity within one week due to oxidation of the thiol group. This loss of activity can only partially be restored by the addition of thiol reagents. Long-term storage can be achieved either by addition of an equimolar amount of mercuric chloride (to yield mercuripapain) or by using 2,2'-dipyridyl disulphide to produce 2-pyridyl papain disulphide. Both of these derivatives are inactive and stable for several months at 4°C. They can be reactivated using thiol reagents such as 2-mercaptoethanol (2ME) or Dithiothreitol (DTT).

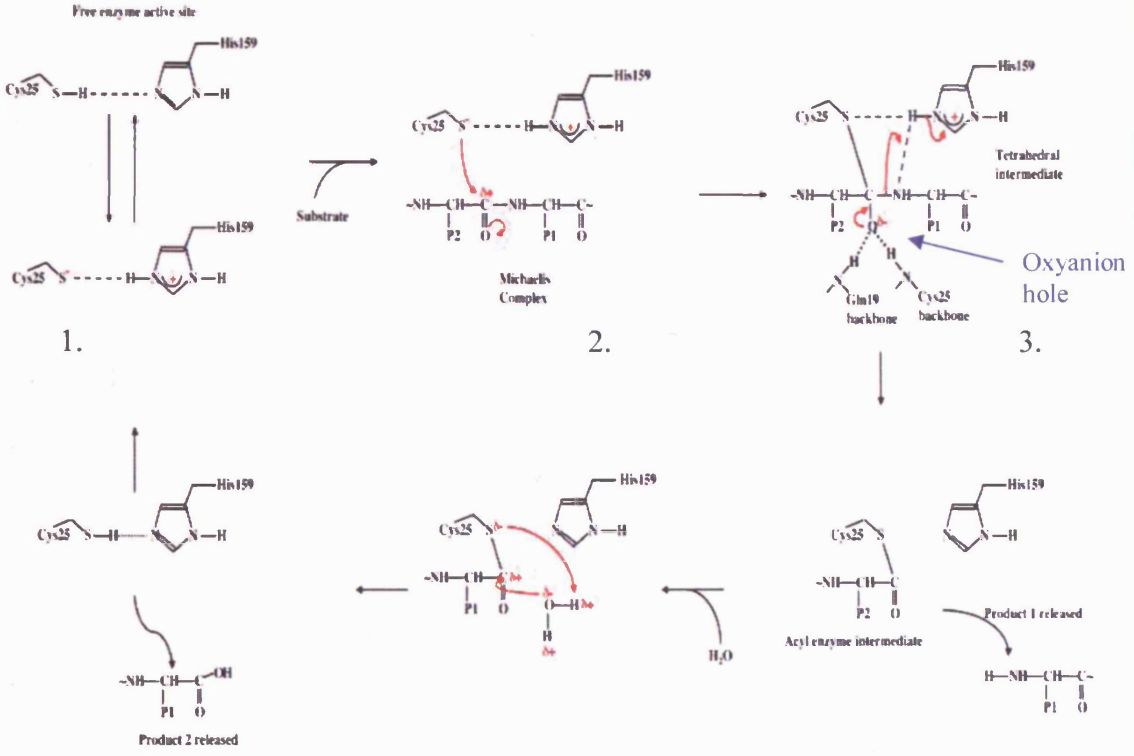
Papain is produced *in vivo* as an inactive precursor or proenzyme – propapain (Brocklehurst and Kierstan 1973, Cohen *et al.*. 1986, Vernet, Tessier *et al.* 1990) which is converted to Papain on activation (Brocklehurst and Kierstan 1973; Vernet, Khouri *et al.* 1991). Recombinant proenzyme has been expressed as soluble fully folded protein in baculovirus cells (Vernet, Tessier *et al.* 1990, Vernet, Khouri *et al.* 1991), and as inclusion bodies (which required refolding) in *E.coli* (Taylor, Pratt *et al.* 1992). In both instances however, the final yields of pure active protein were low (1–2 mg per litre of cell culture). For further information on Papain there are reviews by Baker and Drenth (1987) (Baker and Drenth 1987), Brocklehurst *et al.* (1987) (Brocklehurst, Willenbrock *et al.* 1987), and Storer and Menard (1994 and 1996) (Storer and Menard 1994).

Comparison of HIV-2 PR and Papain mechanisms.

The mechanism of action for HIV-2 PR is dealt with in Chapter 1, whilst that of papain is described briefly below, and in the following references (Klein and Kirsch 1969; Lowe 1970; Brocklehurst and Kierstan 1973; Brocklehurst and Malthouse 1978; Smolarsky 1978; Smolarsky 1980; Brocklehurst, Willenbrock *et al.* 1983; Willenbrock and Brocklehurst 1985; Vernet, Khouri *et al.* 1991; Storer and Menard 1994).

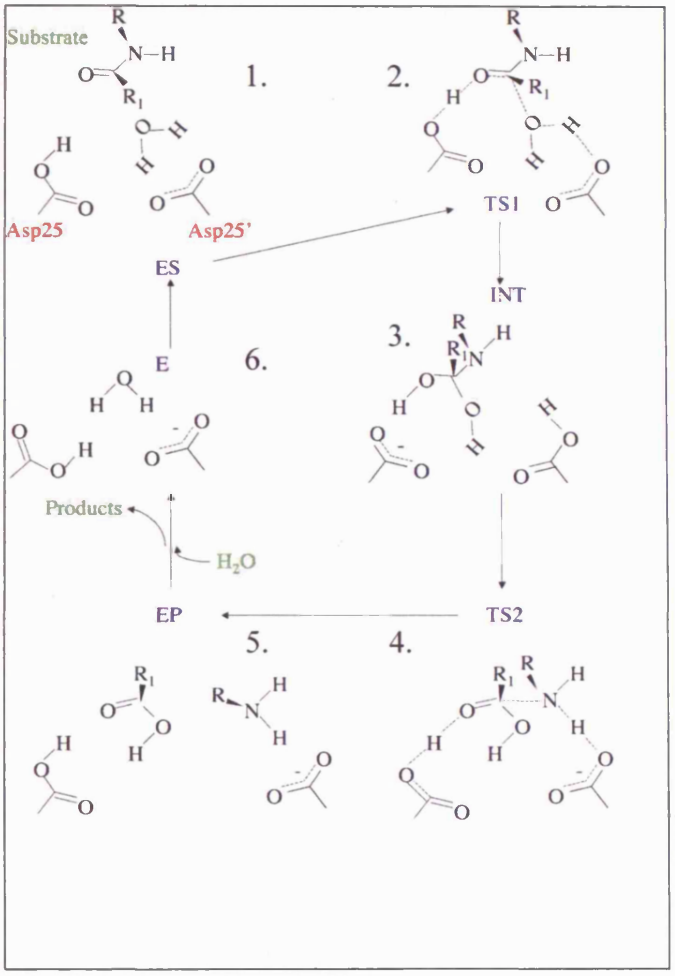
Papain is the best studied of the cysteine family of proteinases, and the cysteine proteinase mechanism, as exemplified by papain, is similar in many respects to that of the serine proteinases. Each mechanism involves an acylation and deacylation step, however, in the cysteine proteinases the attacking nucleophile is the sulphur atom of Cys25 rather than the hydroxyl group of a serine. The proximity of His159 to Cys25 stablises the existence of a thiolate ion (fig. 2) by allowing formation of an ion pair between the Cys25 sulphur and the imidazolium group of His159. Catalysis is possibly also facilitated by the effect of hydrogen bonding between the Asp158 side chain carboxylate and the imidazolium of His159. On binding of substrate to the active site cleft the thiolate ion engages in a nuclephilic attack on the carbonyl carbon of the substrate's scissile peptide bond. A negatively charged tetrahedral transition state results, covalently bound to the enzyme via a thioester bond with Cys25. The transition state is also stabilised by hydrogen bonding between the substrate carbonyl oxygen and the two backbone -NH- groups of enzyme residues Gln19 and Cys25, which together form an 'oxyanion hole' as described by Robertus *et al.* (Robertus, Kraut *et al.* 1972). Collapse of the transition state results in cleavage of the peptide bond; this releases the amino portion of the substrate as the first product (P_1). The acyl portion of the substrate remains bound to the enzyme via Cys25 as a covalent acyl-enzyme intermediate. This completes the acylation step of the mechanism. Deacylation then procedes via enzyme-catalysed attack by water; this generates the acyl product (P_2), and restores the enzyme active site to its free state. A diagramatic representation of the cysteine proteinase mechanism is provided (fig. 4) for comparison with a recent mechanism proposed for the retroviral aspartic proteases (fig. 5).

Figure 4 The mechanism of Papain.



1. Ion pair formation between Cys25 and His159.
2. Thiolate ion nucleophilically attacks substrate carbonyl carbon.
3. Negatively charged tetrahedral transition state forms, with covalent thioester bond between Cys25 sulphur and substrate carbonyl carbon. Hydrogen bonds between substrate carbonyl oxygen and Gln19 / Cys25 backbones form 'Oxyanion hole'.
4. Collapse of transition state, cleavage of peptide bond, release of first product.
5. Covalent acyl intermediate attacked by water. Acyl product 2 released, enzyme active site restored to free state.

Figure 5 General acid-base mechanism for HIV-1 PR (Piana *et al.* 2002).



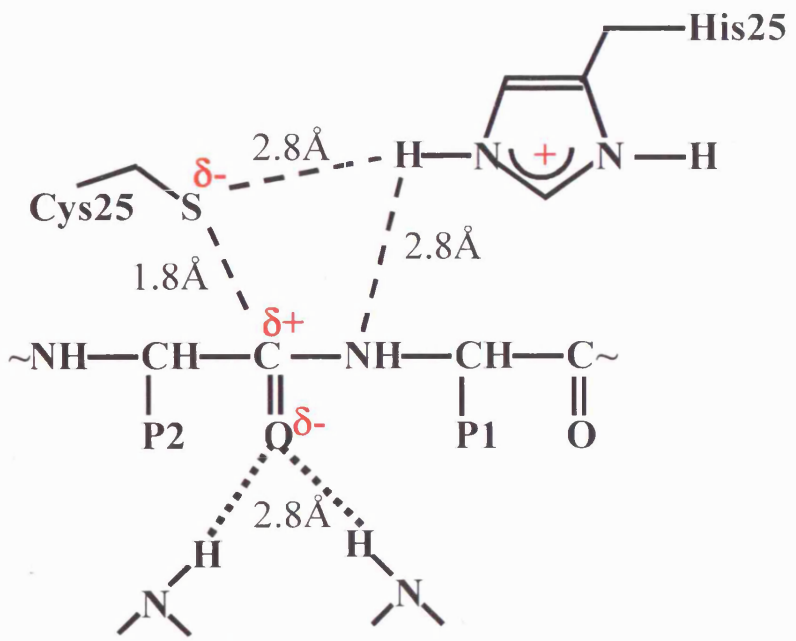
1. Substrate binds, water is displaced, rotations about Asp Cβ-Cα bonds occur, proton transfers occur.
2. Michaelis complex formation, nucleophilic attack by water on substrate carbonyl carbon.
3. Gem-diol tetrahedral intermediate formed.
4. Substrate peptide nitrogen protonated.
5. Substrate peptide bond breaks, products are released, active site regenerated.
6. Free enzyme, catalytic Asps adopt symmetrical monoprotonated state, coordinating a water molecule.

Unlike the papain mechanism there is no formation of a covalent intermediate.

Mutant heterodimer formation.

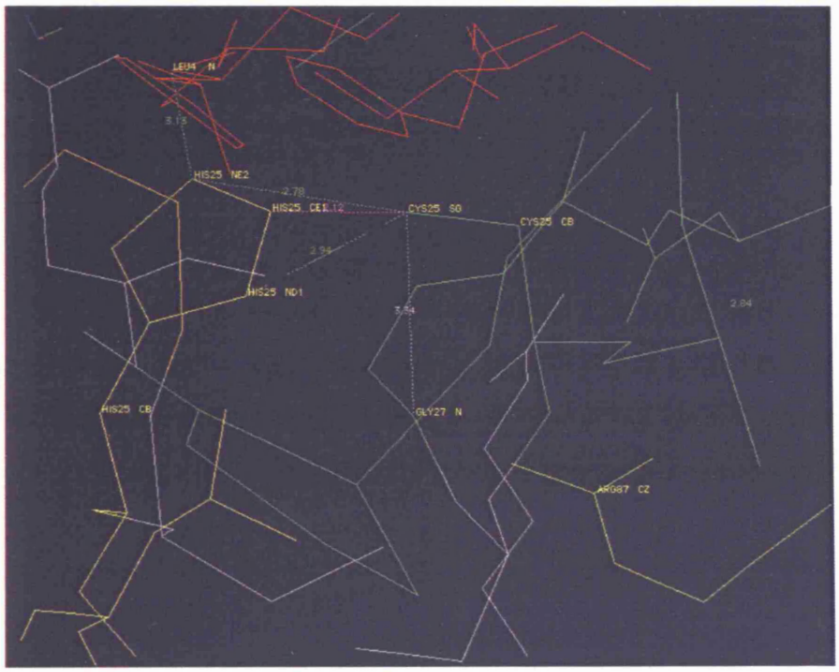
Visual comparison of the respective crystal structures of HIV PR and Papain lead to the proposal that if the two catalytic aspartate residues of HIV-2 PR (D25 on each monomer) were replaced by those of papain (i.e. D25H on one monomer and D25C on the other) the relative spatial orientations of the His and Cys residues in the resultant C25/H25 mutant HIV-2 PR protein would be close to the arrangement of these residues in the native papain active site (fig. 1). Thus it was conceivable that an C25/H25 mutant HIV-2 PR may retain proteolytic activity, despite replacement of the aspartates ordinarily essential for catalysis, because the replacement residues would enable proteolysis by means of a different mechanism - that of Papain. Computer modelling suggested that the replacement catalytic residues could orientate so as to engage a peptide substrate and initiate proteolytic cleavage (figs. 6 and 7).

Figure 6 Catalytic residues of proposed mutant and possible substrate interaction.



His 25 and Cys 25 residues of proposed heterodimeric mutant HIV-2 PR and their predicted interaction with a substrate peptide.

Figure 7 Model of the HIV-2 PR heterodimeric mutant active site with bound substrate.



HIV-2 heterodimer subunits:

D25H = yellow and

D25C = green

Substrate = red

Dotted lines indicate distances (angstroms).

The above figure shows one view of a model of the heterodimeric HIV-2 mutant binding an HIV-2 PR substrate. The distances between the catalytic residues and the substrate approximate to those required for catalysis (ie. those seen in the native papain active site, and illustrated in fig. 6).

The C25/H25 protease thus envisaged would be a heterodimeric active site mutant of HIV-2 PR. The mechanism of action for such a mutant HIV-2 PR would be that of a cysteine protease (papain) rather than that of an aspartic protease (HIV-2 PR), whilst the enzyme's substrate specificity should remain unchanged.

This Chapter describes the production of two homodimeric mutants of HIV-2 PR, D25C (ie C25/C25) and D25H (ie H25/H25) and the attempts to form and detect the activity of a heterodimeric mutant derived from them. It was envisaged that the heterodimeric mutant could be formed by means of monomer exchange between the two homodimeric mutants as a result of dynamic equilibrium between monomeric and dimeric states of the protein. A similar equilibrium has been observed in the native enzyme (Grant, Deckman *et al.* 1992). It was proposed that mixing the two inactive homodimeric mutant forms of HIV-2 PR should produce a heterodimeric mutant detectable by virtue of its activity.

MATERIALS AND METHODS

Three dimensional modelling

Modelling was carried out using Swiss model and Swiss PDBviewer, both available on the internet at http://www.expasy.ch/swissmod/SWISS-MODEL.html.

Mutagenesis reactions

The 'QuickChange' Site Directed Mutagenesis method from Stratagene was used to generate the desired mutants. Primers were designed exactly following the recommendations from Stratagene (see Chapter 2, page 129, Site directed mutagenesis; and the Stratagene QuickChange Manual, p4) and were purchased 'ReadyPure' from PE-Applied Biosystems UK, arriving ready to use. Two primers were required for each mutation (one foward, one reverse) - each the exact complement of the other. The primers (fig. 8) were designated His+ve, His-ve and Cys+ve, Cys-ve. The primers were designed such that they would also introduce a unique AgeI restriction enzyme site, thus enabling clones to be screened for successful mutation by restriction digest.

The plasmid pHIVSET5 was used as the mutagenesis template. This was constructed (fig. 9 i-iii) by insertion of the HIV-2 PR gene as an NcoI/EcoRI PCR product into pRsetB (Invitrogen). The polylinker between the His tag encoding region and the start of the gene was then excised using the flanking NheI and NcoI sites which were cut, filled and religated, maintaining the orginal frame in the process (fig. 9 iv-v). This truncated the construct to remove the enterokinase cleavage site such that the expressed protease would have an N-terminal histidine tag for affinity purification directly before the protein.

The reaction mixtures and PCR cycles were as recommended by Stratagene, and described in Chapter 2, page 129, Site directed mutagenesis.

Figure 8 HIV-2 active site mutagenesis primers.

Asp to His mutagenesis primers.

+ve 5'-ccagtagaagttttgttacacaccggtgctgacgactcaatag-3'

-ve 5'-ctattgagtcgtcagcaccggtgtgtaacaaaacttctactgg-3'

Asp to Cys mutagenesis primers.

+ve 5'-gccagtagaagttttgttatgcaccggtgctgacgactcaatag-3'

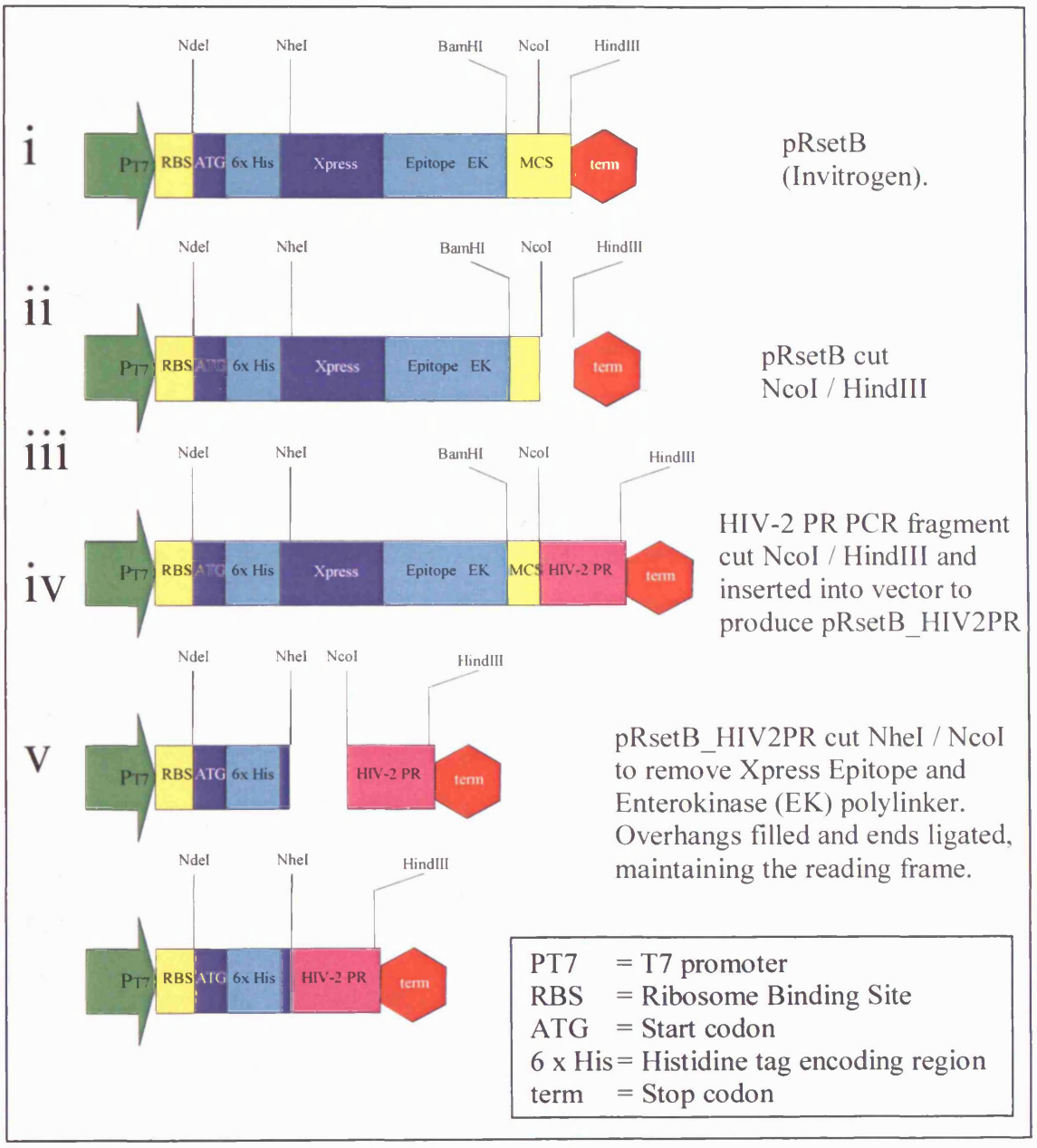
-ve 5'-ctattgagtcgtcagcaccggtgcataacaaaacttctactggc-3'

Blue = DΔH or DΔC Red = AgeI site.

Transformations

Super competent XL1 Blue *E.coli* cells were generated as described in Chapter 2 (See General Materials and Methods, Transformations, The preparation of super competent cells, page 139) and transformed with the completed PCR reaction mixtures described above. The Stratagene 'QuickChange' protocol was followed to effect the transformation, and cells were then spread onto LB agar plates, supplemented with Ampicillin (concentration = 100 $\mu g ml^{-1}$), and incubated overnight at 37°C.

Figure 9 The construction of pHIVset5.



Selection and sequencing of mutants

Colonies resulting from the above transformation were used to generate plasmid minipreps (as described in Chapter 2, page 124, Preparation of DNA, Small scale or 'Minipreps'). The minprep DNA was then digested with both AgeI and HindIII as a diagnostic test to identify positive plasmids. Only clones containing the desired mutations cut with both enzymes to yield a linearised DNA fragment of approximately 350 bp. Midiprep DNA (Chapter 2, page 125, General Materials and Methods, Preparation of DNA, Large scale 'Midipreps') was prepared from the positive clones (ie. those containing both sites), and was used for sequencing of the mutant genes. ABI sequencing was carried out by Cambridge biosciences Ltd (see page 133, and the List of suppliers, Appendix, page 325).

Expression and purification of the mutant homodimers

Each of the two mutants was expressed and purified individually in the form of mutant homodimers (i.e. H25/H25 and C25/C25), using BL21 (DE3) pLysS as the host strain. Cultures were innoculated with freshly transformed cells and grown to an OD_{280} of 0.5 at 37°C, then allowed to cool to 16°C and induced by addition of IPTG to a final concentration of 0.1 mM. The cultures remained incubating with shaking at 16°C overnight, and were then harvested by centrifugation. Cell pellets were resuspended in Buffer J (table 1) and stored at -70°C pending to use.

Table 1 Buffer J.

50 mM Tris pH 8.0, 100 mM NaCl, 1 mM PMSF, 0.01% Sodium azide

Frozen cells were thawed and lysed by sonication as required. Cell lysates were clarified by ultra centrifugation at an RCF of 141000 g for 1 hour at 4°C. The mutant protease was isolated from the supernatant fraction by Talon IMAC chromatography (see Materials & Methods, page 159) following the manufacturers (Clonetech) recommended protocol. The eluted protease solution was then supplemented by addition of EDTA and DTT, each to a concentration of 1 mM, and reduced in volume using an Amicon pressure cell prior to further purification by SE chromatography using an XK26/60 Superdex S200 column (Pharmacia). The final protease solution was dialysed into Buffer G (see Chapter 4, table 1, page 207), and its volume reduced by means of Amicon pressure cell, then Centricon 3, until a concentration of approximately 10 $mgml^{-1}$ was achieved.

Characterisation of the mutant homodimers.

PAGE and Western blot analysis.

Each of the two purified mutant proteases were analysed by SDS-PAGE and Western blots using anti-HIV-2 PR, and anti-histidine tag primary antibodies to ensure that they were the histidine tagged mutants of HIV-2 PR.

Solubility screen.

The mutant proteins were observed to form light precipitate during storage. A solubility screen was therefore carried out in order to establish solubility characteristics of the mutants compared with the native enzyme. The screen was carried out manually, but in principle used the method later described in Hinks *et al.* 2003 (Hinks, Roe *et al.* 2003). A buffer grid of 1 µl drops varying in pH (50 mM Phosphate pH 4 - 8) and NaCl concentration (0 - 1000 mM in 200 mM increments) was set up under paraffin oil in a 48 well microtitre plate. An equivalent volume of the protein to be analysed was then dispensed into each well and mixed with the contents. The droplets formed were then observed using a light microscope (set to dark field illumination) to determine the extent of precipitation, if any, in each well. In this way the optimal and pessimal conditions for protein solubility could be established in terms of the pH and salt content of its buffer. An initial observation was made immediately after dispensing the protein, and a subsequent observation was made after 24 hours at 14°C.

Activity assay.

Chromogenic substrate VII (Bachem) was used to assay each mutant for possible proteolytic activity (method as described in Chapter 2, General Materials and Methods, Protease assays, Kinetic analysis of proteases, page 164).

Formation and detection of the mutant heterodimer

The methods by which formation of this heterodimeric mutant was attempted are described below. In each case, activity was assayed by means of chromogenic substrate (See Chapter 2, General Materials and Methods, Protease Assays, Kinetic analysis of proteases, page 164). The SPA assay would have been more suitable for the detection of heterodimer formation. However, this option was no longer available in our laboratory at the time of experimentation.

Simple mix method

The pure mutant homodimers (in solution at 2 mg/ml) were mixed in approximately equimolar amounts, and 10 µl of the mixture was added to 1 ml of the activity buffer, at 37°C in the presence of chromogenic substrate (a baseline absorbance for buffer and substrate at 300 nm having previously been established). A reduction in absorbance at 300 nm was indicative of substrate cleavage. The presence of heterodimer with the desired papain active site would thus be detected as proteolytic activity.

Initial measurements were carried out over a period of 180 seconds, and subsequent readings were taken after 20 hours. Since it was anticipated that heterodimer formation might be a slow process, the long incubation was intended to detect the cumulative effect of any substrate cleavage that may occur.

Refolding method

Each of the homodimers was separately dialysed into buffer G supplemented with 8 M urea to produce two solutions containing the denatured mutant monomers.

Approximately equivalent molar amounts of the two denatured monomer solutions (each 2 mg/ml) were mixed, and the mixture was then dialysed back into non-denaturing buffer G (i.e. not supplemented with urea) to refold the mixed proteins. The refolded proteins were then assayed for proteolytic activity as described above.

Assay conditions

Activity assays were carried out at conditions optimal for both HIV-2 PR, and papain. A pH range between 4 and 8 was investigated, in conjunction with salt concentrations of 0 and 1 M NaCl. Sodium acetate (100 mM) buffer was used for solutions at pH 4 and 5, whilst 100 mM phosphate was used to buffer solutions between pH 6 and 8 inclusive. DTT was incorporated into all buffers at a final concentration of 1 mM, and freshly prepared protein was used for all the assays. The effect of temperature was not investigated, and 37°C was used in all cases.

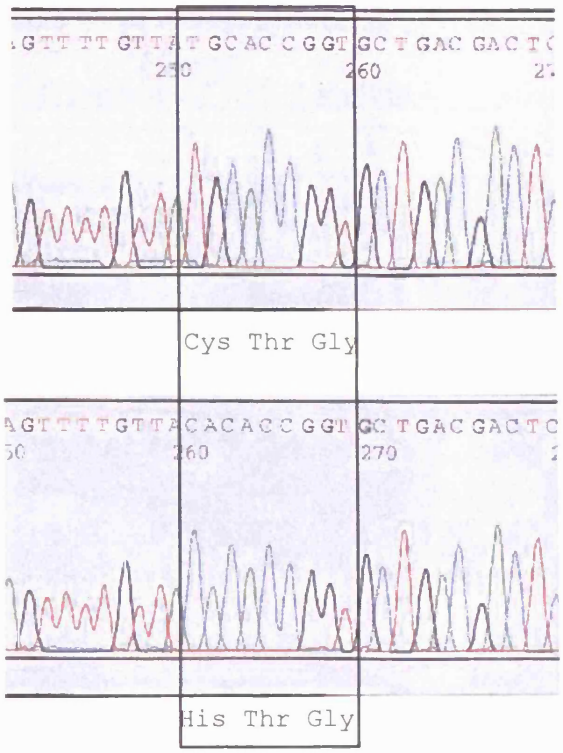
RESULTS AND DISCUSSION

Mutagenesis

A number of mutagenesis systems were considered and attempted without success during the preliminary stages of this work, among them the Recombinant Circle method (Jones and Howard 1990) and the 'Altered Sites' mutagenesis system, by Promega. Finally the 'QuickChange Site Directed Mutagenesis' method was employed following several personal recommendations, supporting the assertion by Stratagene that it was indeed a "reliable means to conduct site-directed mutagenesis" (QuickChange manual, Stratagene). Generally the Stratagene protocol was observed throughout (as described in Chapter 2, Site Directed Mutagenesis, page 129) however, there was no necessity to purchase the kit itself. All the enzymes required, such as Pfu DNA polymerase (selected for its fidelity and proof reading ability) and DpnI restriction endonuclease (used to remove the original template DNA), could be bought more cheaply elsewhere (see Chapter 2, page 123, Suppliers of reagents and commercial enzymes; and the List of suppliers, Appendix, page 325). The Stratagene mutagenesis protocol utilises only 16 amplification cycles in order to minimise non-specific mutations, consequently the PCR product yield is low. This necessitates the use of super competent *E.coli* cells (XL1-Blue), and these were produced following the method previously described (see Chapter 2, page 139, Transformations, Preparation of super competent cells). Potential mutants were screened by AgeI digestion followed by ABI sequencing of clones positive for the restriction site (fig. 10).

Figure 10 Sequencing confirmed the successful mutagenesis of HIV-2 PR.

D25ΔC25



D25ΔH25

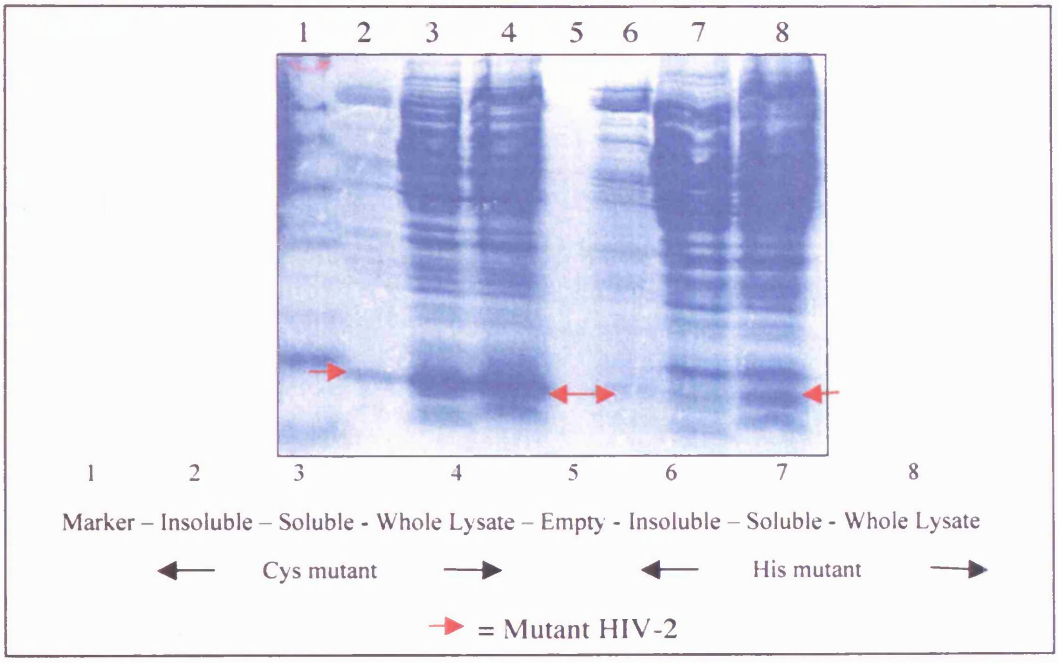
ABI sequencing confirmed the successful mutagenesis of HIV-2 PR.

The native active site Asp 25 residue was changed to Cys in the D25ΔC25 mutant (top), and to His in the D25ΔH25 mutant (bottom).

Expression and purification

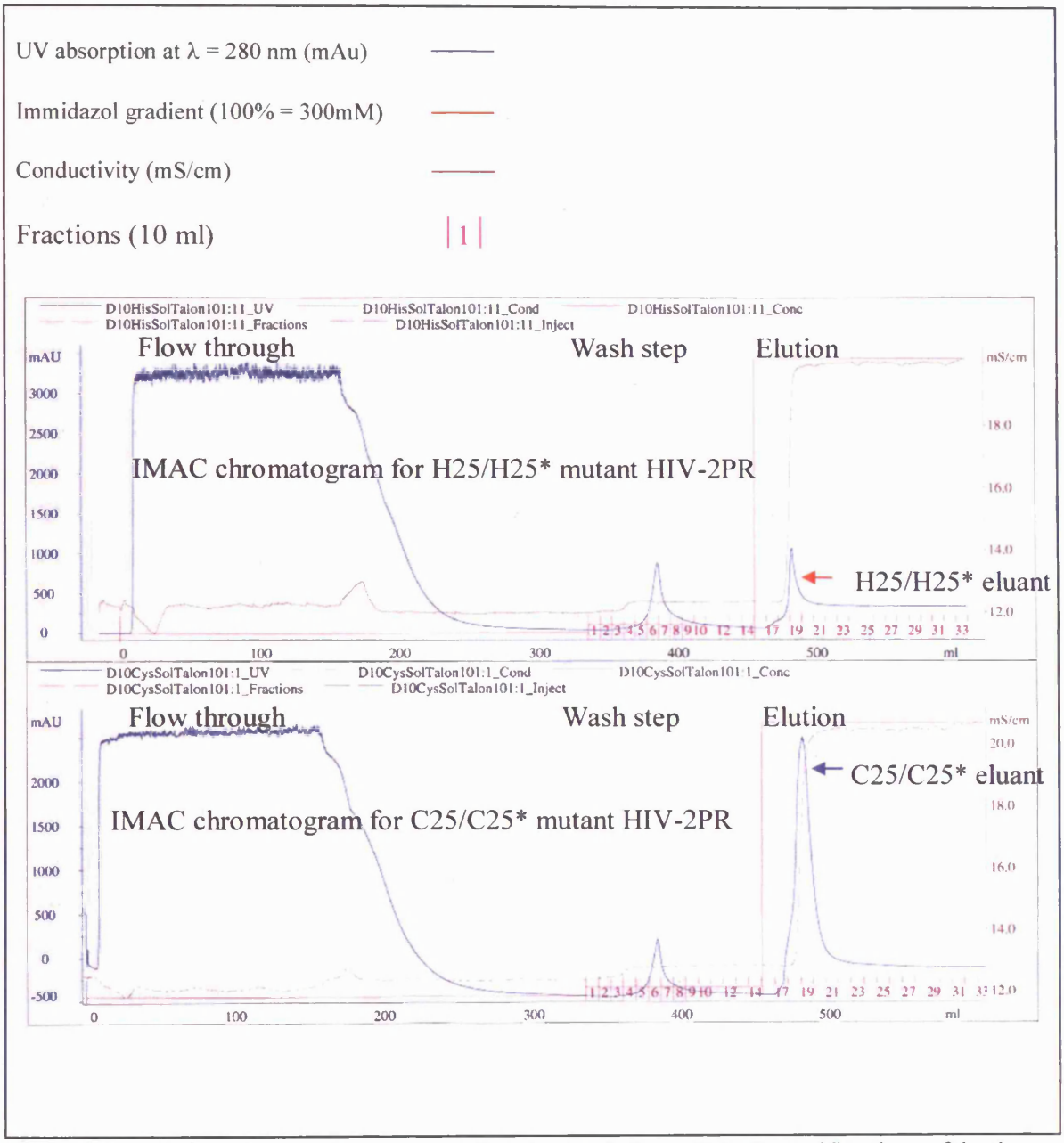
Here 0.1 mM IPTG was used at an induction temperature of 16°C, ensuring a slow expression rate. This was observed to improve yields of the soluble protein, presumably reducing the rate of expression allowed the protein time to fold correctly. Figure 11 shows expression of the mutant proteases as visualised by SDS PAGE.

Figure 11 SDS-PAGE analysis of HIV-2 mutant expression in *E.coli.*



In both cases it was observed that there was a significant improvement in yield over native HIV-2 PR preparations, however it was also noted that the expression levels were higher for the C25/C25 mutant than for the H25/H25 mutant. This was particularly apparent when the results from the IMAC purification step for the two mutants were compared (fig. 12). The purification process for each mutant was the same (IMAC followed by Size Exclusion), and consequently the SDS-PAGE analyses of each step are shown for the Cys mutant only (figs. 14 and 16).

Figure 12 Comparison of IMAC data for the HIV-2 PR mutants.

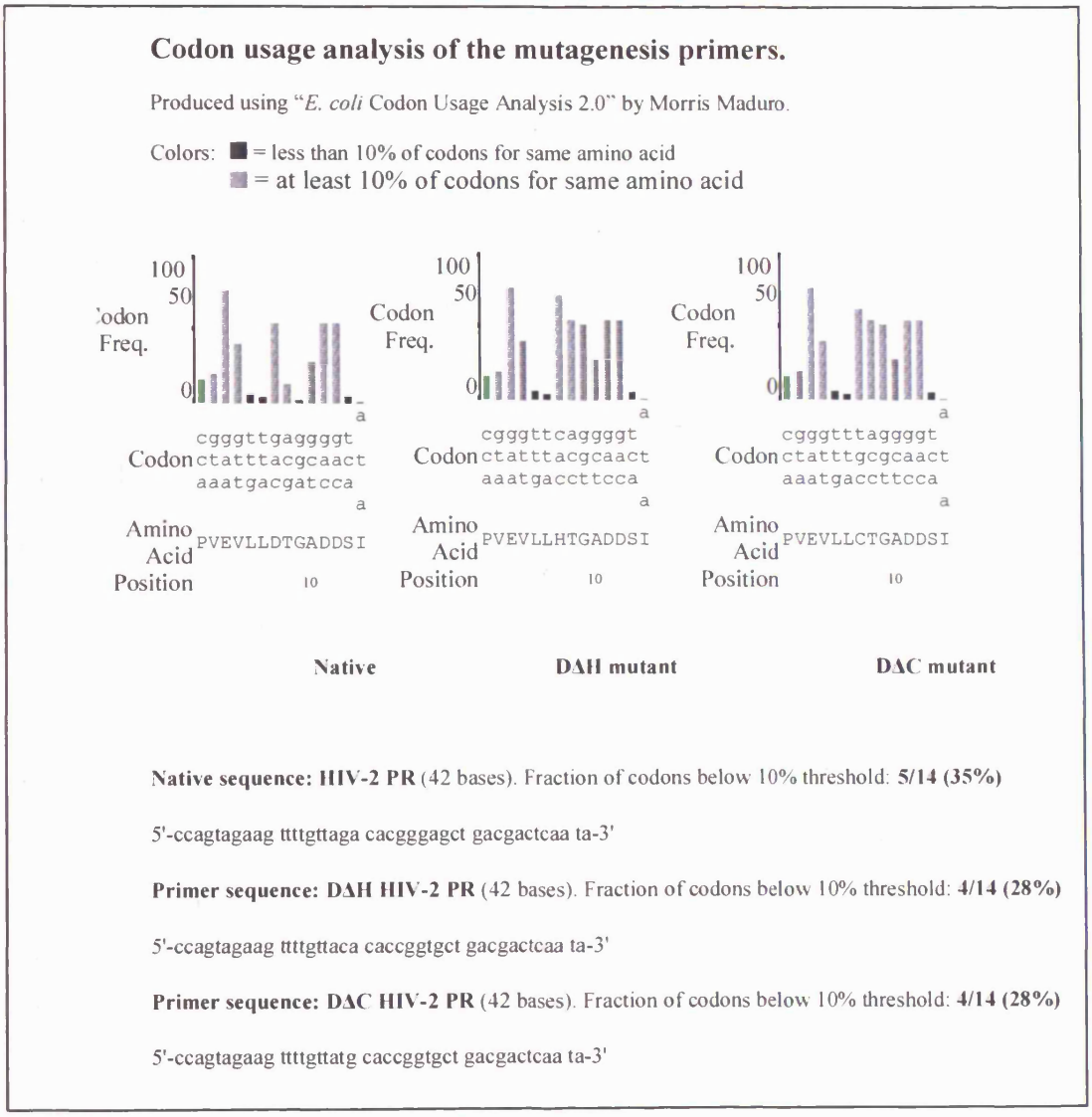


The above figure shows chromatographs obtained from IMAC purification of both H25/H25 and C25/C25 mutants (top and bottom respectively). Cell cultures were prepared in the same way in both cases, equivalent volumes of cell lysate were loaded and the identical column and instrumentation was used for both runs. A comparison of the chromatographs shows that twice as much protein from the C25/C25 lysate bound to the Talon resin compared with the H25/H25 lysate. This reflects the PAGE results, and suggests that in this system expression of the C25/C25 mutant is favoured over expression of the H25/H25 mutant.

It has for some time been accepted that retroviral proteases are - by virtue of their proteolytic activity - toxic to cells in which they are expressed, and that this toxicity is responsible for the poor yields observed (Rangwala, Finn *et al.* 1992; Stebbins and Debouck 1994; Gustafson, Junger *et al.* 1995; Wan and Loh 1995). The improved expression observed here might therefore be attributable to the disruption of proteolytic activity by mutagenesis.

However, the difference in expression levels between the constructs is harder to explain, and suggests an additional factor or factors effecting protease expression in *E.coli.* It may be that toxicity is a major cause of poor protease yields, but the results presented here, and in Chapter 6 suggest that this is probably not the only influence. The presence of codons which are rare in *E.coli* would also explain poor expression levels (this is discussed at length in Chapter 6). The comparison of codon usage between the native and mutant proteases (fig. 13) illustrates that the HIV-2 PR active site Glycine is encoded by a rare codon (GGA) in the native enzyme that has been optimised in the mutants. Similarly the less frequent Threonine codon (ACG) also found in the native enzyme has also been optimised, raising its frequency from approximately 10% to about 40%. The native protein contains five rare Glycine encoding codons (see Chapter 6 fig. 20) and they are the least frequent of the protein's codons. The elimination of one such rare codon, plus the optimisation of the Threonine codon may be expected to have a beneficial effect on expression levels of the HIV-2 PR mutants. This combined with the elimination of proteolytic activity would explain the improved expression of the mutants. However, since the mutants differ by only a single amino acid, and since the codon selected to encode for the new amino acid was optimised for *E.coli* expression, this is not a likely explaination for the difference in expression between the two mutants.

Figure 13 Comparison of codon usage for HIV-2 native and mutant PR.



Native sequence: HIV-2 PR (42 bases). Fraction of codons below 10% threshold: **5/14 (35%)**

5'-ccagtagaag ttttgttaga cacgggagct gacgactcaa ta-3'

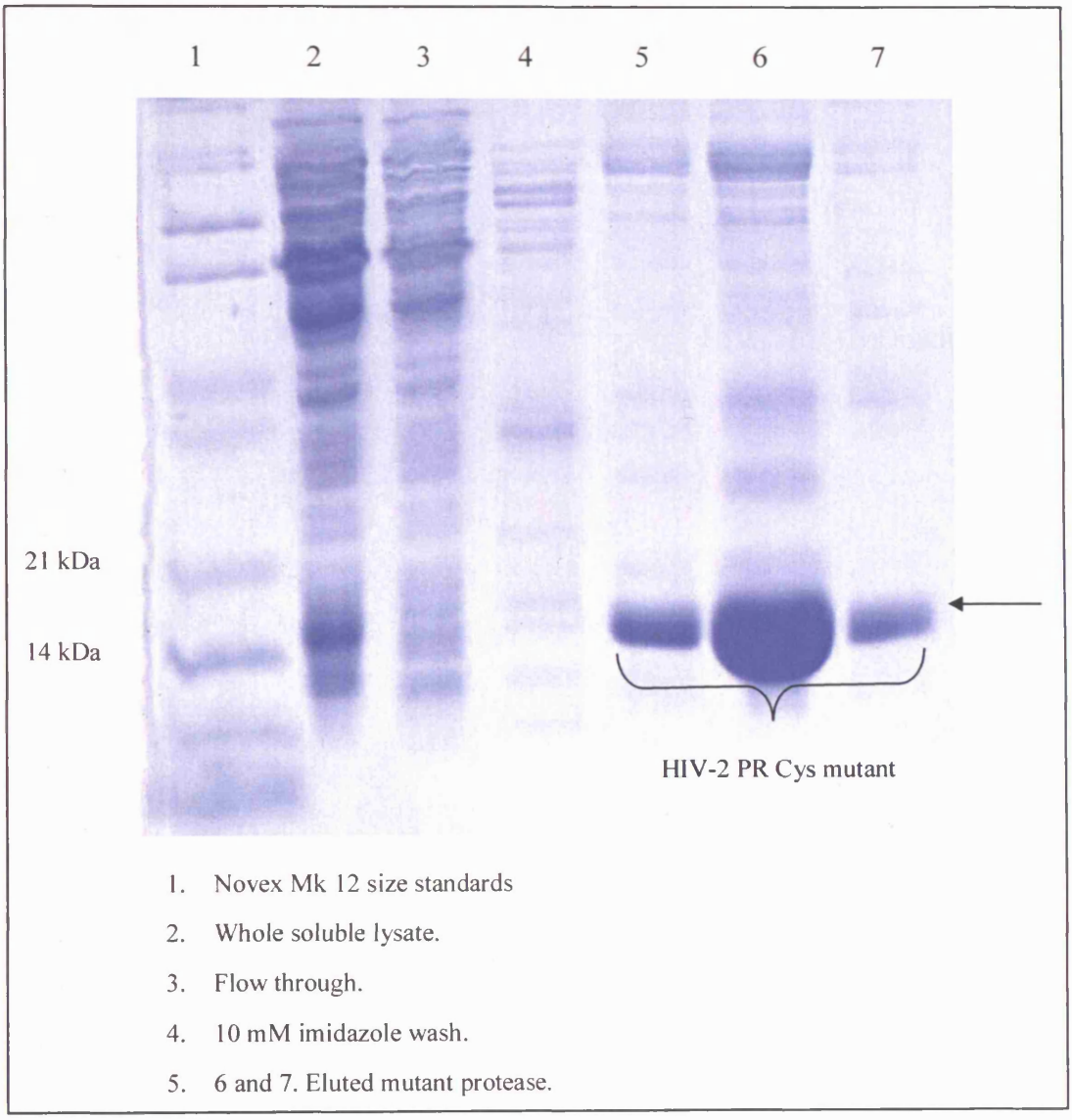
Primer sequence: DΔH HIV-2 PR (42 bases). Fraction of codons below 10% threshold: **4/14 (28%)**

5'-ccagtagaag ttttgttaca caccggtgct gacgactcaa ta-3'

Primer sequence: DΔC HIV-2 PR (42 bases). Fraction of codons below 10% threshold: **4/14 (28%)**

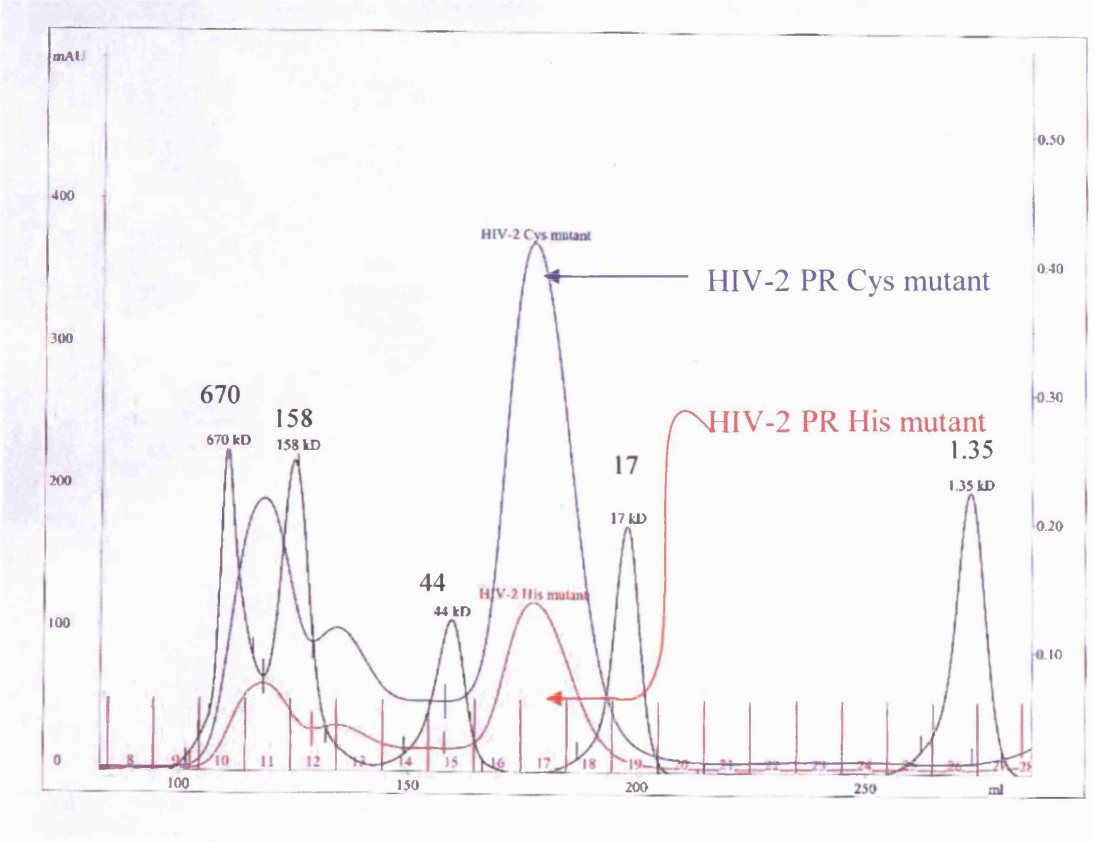
5'-ccagtagaag ttttgttatg caccggtgct gacgactcaa ta-3'

Figure 14 SDS-PAGE of IMAC fractions for HIV-2 PR Cys mutant.



1. Novex Mk 12 size standards
2. Whole soluble lysate.
3. Flow through.
4. 10 mM imidazole wash.
5. 6 and 7. Eluted mutant protease.

Figure 15 Chromatographs following size exclusion chromatography.

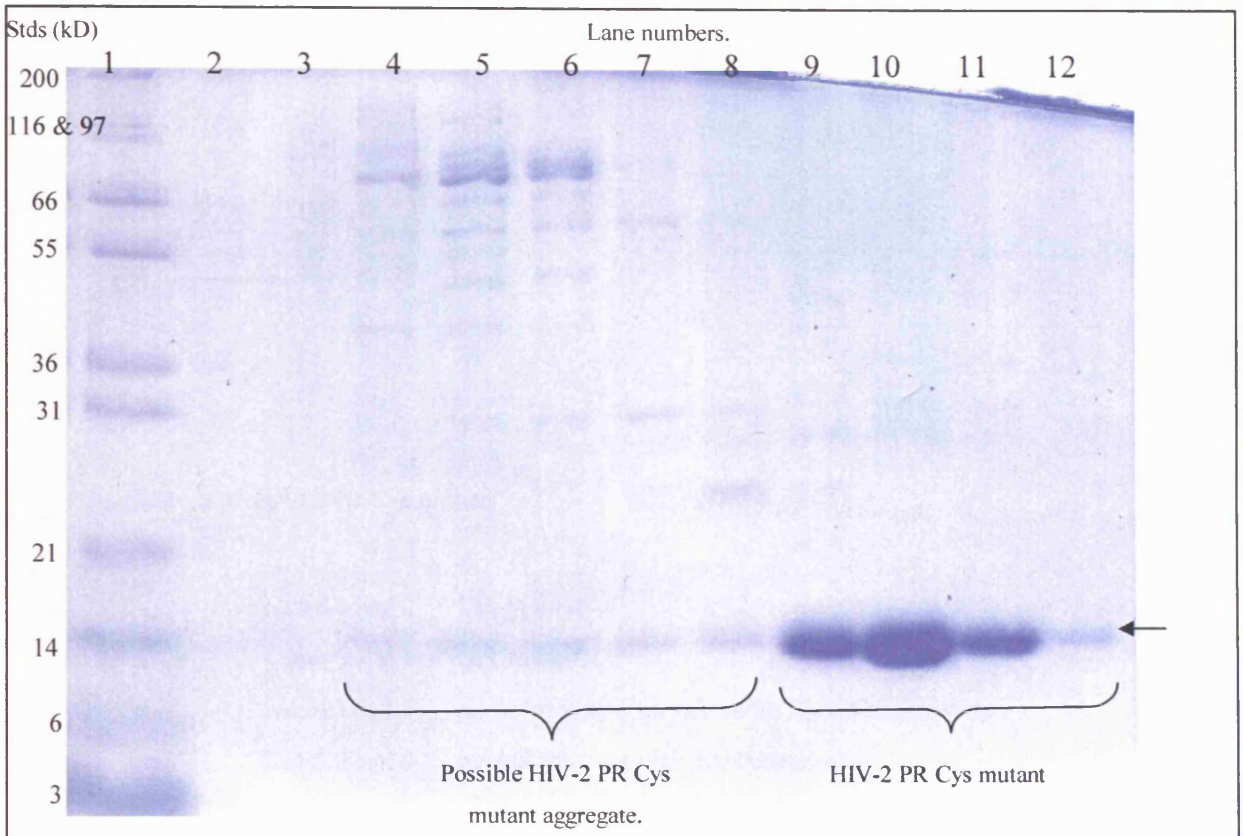


Chromatographs obtained using a Superdex S200 Column, and superimposed.

Black	Size standards (values in kDa)
Blue	HIV-2 PR Cys mutant
Red	HIV-2 PR His mutant

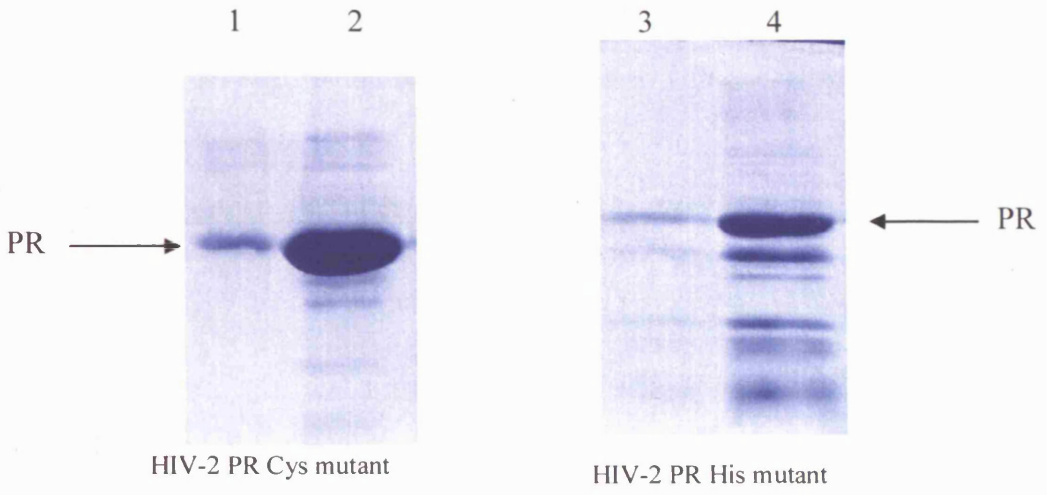
Each mutant eluted with a retention volume equivalent to a protein of approximately 22 kD, suggesting that they were both dimers in solution.

Figure 16 SDS-PAGE of fractions following size exclusion chromatography.



HIV-2 PR mutants eluted from the Superdex S200 column with a retention volume consistent with a protein of approximately 22 kD, and ran at approximately 11 kD on a denaturing gel (lanes 9-12), suggesting that they were dimeric. There was also some evidence of possible aggregation, since trace amounts of protease were visible in earlier fractions (lanes 4-8). This was especially true of the His mutant.

Figure 17 SDS-PAGE of purified and concentrated mutant HIV-2 PRs.



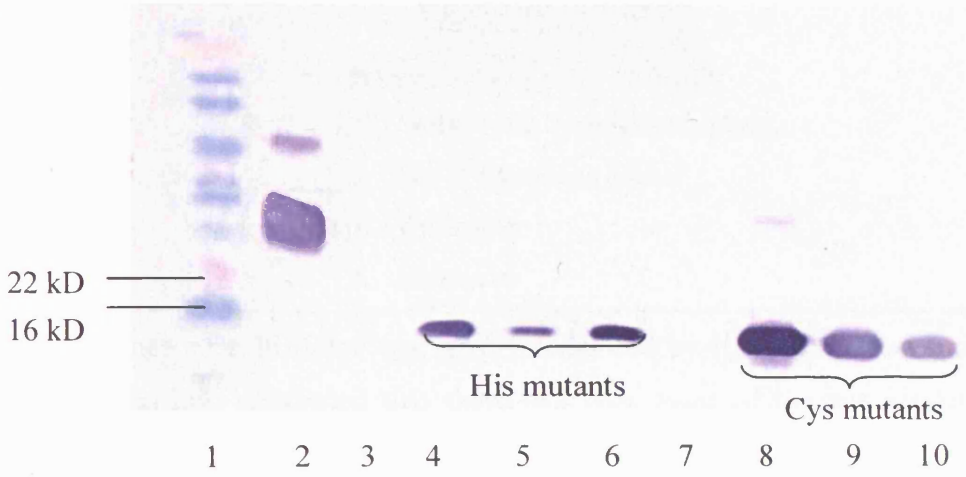
Lanes 1 and 2	Cys mutant
Lanes 3 and 4	His mutant
Lanes 1 and 3	pooled fractions following final purification step
Lanes 2 and 4	pooled fractions after concentration

The His mutant appears less pure once concentrated than the Cys mutant. This is primarily because the degree of concentration for the His mutant was two-fold greater than that of the Cys mutant in order to achieve an approximately equivalent final concentration. As a result the contaminants present in both appear more prevalent in the His mutant preparation.

Characterisation of the mutant homodimers.

Each of the two mutants ran on a coomassie stained SDS-PAGE gel (figs. 14, 16, 17) at a size of approximately 11 kD. When analysed by Western blot both mutants cross-reacted with anti-HIV-2 PR antibody (fig. 19), and with anti-Histidine tag antibody (fig. 18).

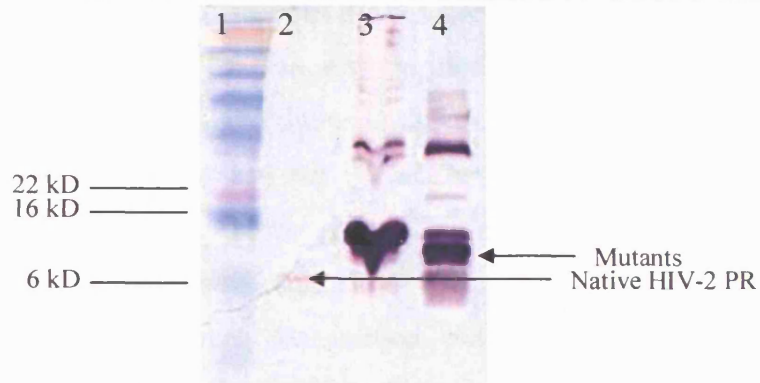
Figure 18 Western blot of Histidine tagged mutant HIV-2 proteases.



Primary antibody = anti His tag

1. Seeblue plus 2 prestained markers
2. Positive control His tagged protein
3. Negative control E.coli lysate
4. DΔH HIV-2 PR (whole lysate)
5. DΔH HIV-2 PR (soluble fraction)
6. DΔH HIV-2 PR (insoluble fraction)
7. Unused lane
8. DΔC HIV-2 PR (whole lysate)
9. DΔC HIV-2 PR (soluble fraction)
10. DΔC HIV-2 PR (insoluble fraction)

Figure 19 Mutants cross react with HIV-2 PR specific antibody.

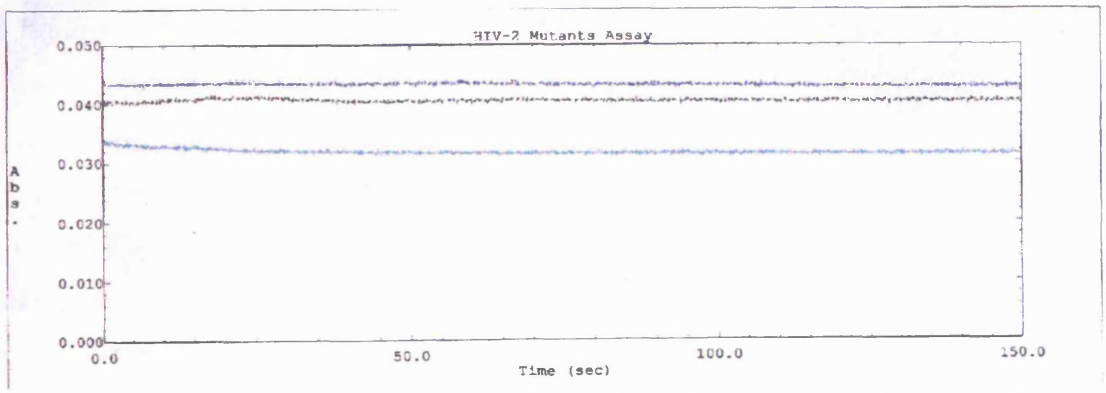


Primary antibody = anti-HIV-2 PR

1. SeeBlue plus 2 prestained markers
2. HIV-2 PR positive control
3. Cys mutant
4. His mutant

The presence of a histidine tag, as demonstrated by the purification procedure and western blotting, confirmed that there had been none of the autocatalytic cleavage characteristic of active retroviral proteases at the N-terminal. Neither mutant exhibited any proteolytic activity when assayed in isolation, even when present in high concentration, confirming that all HIV-2 PR activity had been knocked out (illustrated in fig. 20).

Figure 20 Chromogenic assay of homodimeric mutants illustrating inactivity.



Blue (top)	= Cys mutant	A constant absorbance with time in all
Black (middle)	= His mutant	cases indicated inactivity.
Light blue (bottom)	= substrate alone	

HIV-2 PR mutant homodimers are less soluble than the native PR.

The mutants were overexpressed in both the soluble and insoluble fractions, though only protein from the soluble fraction was purified for analysis. However, during storage (or sometimes as a result of dialysis) the mutant proteins had a tendancy to precipitate out of solution. This had not been observed with the native HIV-2 PR. Oxidation of the active site cysteine may be implicated in precipitation of the C25/C25 mutant homodimer, but could not explain precipitation of the H25/H25 homodimer. A solubility screen was carried out (figs. 21 and 22), and this demonstrated that the mutants were significantly less soluble at low pH conditions (ie. between pH 4 and 6) than the native protease. This range incorporated the pH conditions under which native enzyme activity would normally be seen, and precipitation was even observed in the HIV-2 PR activity buffer routinely used for native assays.

It was also noted that neither of the mutant homodimers could be crystallised using the optimal conditions for the native HIV-2 PR. The single amino acid change buried within the PR active site was not expected to have very much effect on solubility or crystallisation, and yet it did. The results may indicate a change in the protease conformation following mutation, or perhaps that the mutants failed to fold correctly thus exposing hydrophobic residues and resulting in reduced solubility.

HIV-1 PR has been investigated for the presence of cooperative folding units (Wallqvist, Smythers *et al.* 1998), defined as "structural units that exhibit a relatively stronger protection against unfolding than do other parts of a given molecule" (Wallqvist, Smythers *et al.* 1998). Wallqvist *et al.* have identified a folding core within HIV-1 PR consisting of the following spatially close sets of residues 22-32, 74-78, and 84-91. Mutations seen in HIV-1 PR, occuring either naturally or as a result of induction by drug therapy, are predominantly (if not exclusively) within regions not essential to the structural stability of the protein (ie. not within the folding core regions). The consequence of mutations within the folding core regions would be structural instability and a resultant loss of activity, thus they could not persist *in vivo*. It is reasonable to expect that similar folding core regions are present in other retroviral proteases, such as HIV-2 PR.

Figure 21 Solubility screen results at time zero.

Solubility Screen

Time Zero

Native HIV-2 PR

[NaCl] mM

pH	0	200	400	600	800	1000
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

D25H mutant

[NaCl] mM

pH	0	200	400	600	800	1000
4	0	0	0	1	2	3
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

D25C mutant

[NaCl] mM

pH	0	200	400	600	800	1000
4	0	0	0	1	2	3
5	3	3	3	3	3	3
6	3	2	1	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

Key

0	Clear drop
1	Feint precipitation
2	Very light precipitation
3	Light precipitation
4	Precipitate
5	Heavy precipitate
6	Very heavy precipitate

Figure 22 Solubility screen results after 24 hours.

Solubility Screen

24 hours

Native HIV-2 PR

pH	[NaCl] mM					
	0	200	400	600	800	1000
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

D25H mutant

pH	[NaCl] mM					
	0	200	400	600	800	1000
4	3	3	3	4	4	4
5	1	1	1	1	1	1
6	0	0	0	0	0	1
7	0	0	0	0	0	0
8	0	0	0	0	0	1

D25C mutant

pH	[NaCl] mM					
	0	200	400	600	800	1000
4	2	2	2	2	2	2
5	2	2	2	2	2	2
6	2	2	1	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0

Key

0	Clear drop
1	Feint precipitation
2	Very light precipitation
3	Light precipitation
4	Precipitate
5	Heavy precipitate
6	Very heavy precipitate

The active site residues are numbers 25, 26 and 27 (DTG in the native HIV-1 and 2 proteases, or in the HIV-2 PR mutants HTG and CTG respectively), which puts them within one of the folding core regions. It is possible therefore, that in creation of the active site mutants an HIV-2 PR cooperative folding unit has been disrupted, rendering the mutant proteases less likely to fold correctly. This may provide an explaination for the relatively high levels of mutant protease observed in the insoluble fraction of the cell lysates. It might also explain their reduced solubility and tendancy to precipitate from solution. The analysis by SDS-PAGE of fractions eluted following size exclusion chromatography suggested the presence of some aggregated protein; however, the bulk of soluble protein appeared to be dimeric. It therefore appears that a large proportion of the mutant protein was able to fold and dimerise, but that it was less stable than the native protein and with time the disrupton to the folding core caused further aggregate to form resulting in the precipitation observed.

Localised disruption within the folding core could very likely explain the lack of activity seen on attempted formation of the heterodimeric C25/H25 mutant protease. Minor local misfolding might prevent the anticipated positioning of Histidine and Cysteine relative to each other within the heterodimer, without necessarily having a gross effect on the protein's folding as a whole. Consequently mutant monomer subunits would still be able to dimerise, but the intended formation of a papain active site would not occur. Indeed the formation of the active site cleft may also be disrupted with the result that substrates cannot bind.

It is also possible that such disruption of the 22-32 folding core could prevent the mutants adopting the native HIV-2 PR conformation altogether. In this event dimerisation would not occur, and instead a non-functional, misfolded peptide or aggregate would result. The size exclusion chromatography results showed some evidence of aggregation, suggesting a gross effect on folding (and hence on dimerisation) did result from mutagenesis a proportion of the time. However, since the results also demonstrated the presence of dimer, it seems that predominantly any misfolding was sufficiently slight and localised that gross disruption did not occur and dimerisation remained possible.

However, further experiments would be required to determine whether the mutants are misfolded. Analysis by means of Circular Dichroism (CD) would be desirable,

since this would allow investigation of the mutant protein secondary structure. CD spectra for each homodimer could be compared with spectra for the native enzyme to establish whether or not the mutants were folded protein or misfolded soluble aggregates. Alternatively 1D NMR spectra could be used for the same purpose.

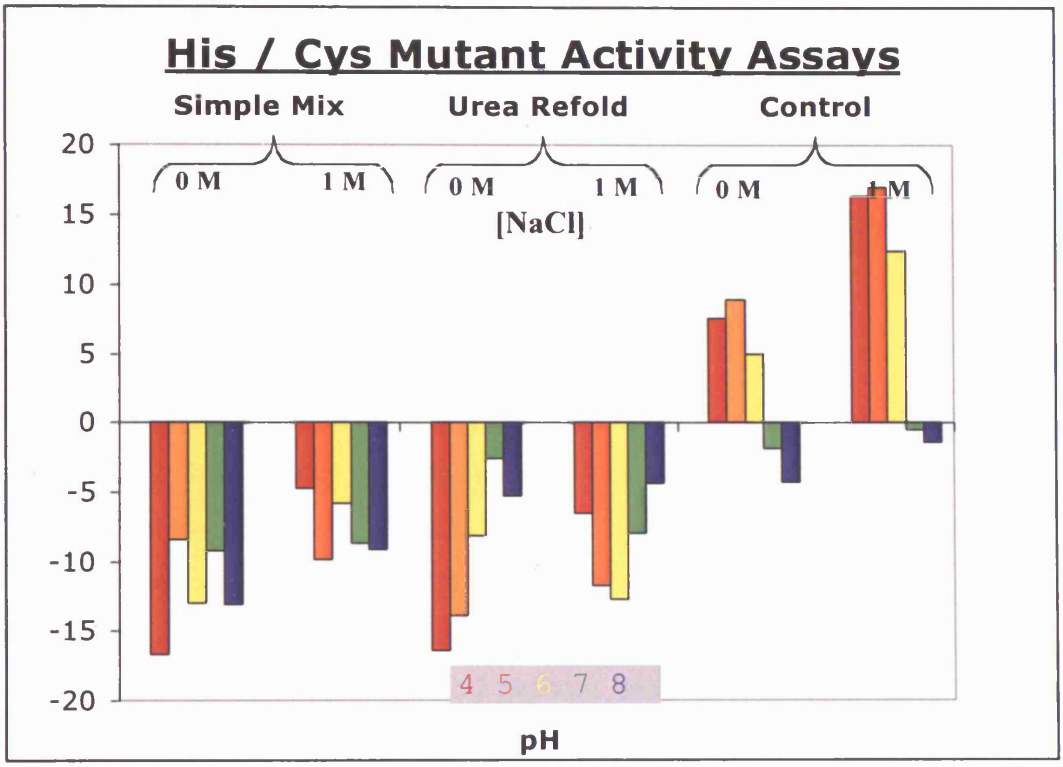
Formation and characterisation of the mutant heterodimer

The mutagenesis resulted in only a single amino acid change in a region not related to dimerisation. Consequently it was anticipated that the two mutant proteins would be expressed in homodimeric form, just as observed for the native protein. Thus for the D25ΔC25 mutant the expressed protein would be a homodimer of C25/C25, while the D25ΔH25 mutant would be expressed as an H25/H25 homodimer. Neither form of mutant homodimer was expected to possess activity. However, in solution the folded dimeric protein probably exists in a dynamic equilibrium with unfolded monomeric protein (Grant, Deckman *et al.* 1992) consequently in a mixed solution of the two mutant homodimers an exchange of monomers should occur with time. This would result in the formation of some C25/H25 heterodimer. This phenomenon of monomer exchange has been exploited to inhibit HIV PR activity via the addition of a defective monomer to a protease solution (Babé and Craik 1991; Babé, Pichuantes *et al.* 1991; McPhee, Good *et al.* 1996). Any C25/H25 heterodimer formed would have the catalytic residues of papain in its active site and may thus possess detectable proteolytic activity.

Assay results for simple mix and refolding methods.

Assays were carried out (as described in Chapter 2, page 164) both in HIV-2 PR activity buffer, and in buffers over a pH range from 4 to 8 and with NaCl concentrations of 0, and 1000 mM. The simple mix and refolding methods were both investigated. No activity (as indicated by a reduction in absorbance with time at λ = 300 nm) was seen under any of these conditions during the first 60 - 180 seconds following mixing. Figure 23 shows the result obtained using HIV-2 PR activity buffer, and it can be seen that there was no reduction in absorbance throughout the period of the assay. This was typical of all the conditions tested, and indicates that there was little or no protease activity present. In order to detect any trace activity that may have been present, and also to allow time for heterodimerisation to occur, the samples were incubated at 37°C overnight and the absorbance at λ = 300 nm was recorded for each assay mixture both before and after incubation. Protease activity would have resulted in a decrease in absorbance over this time, however, no such decrease was seen under any of the conditions tested (fig. 23 and Appendix, tables 6-8, pages 348-351). During the refolding process homo and hetero dimerisation was expected to occur. Given an equimolar mixture of the homodimers, and given that homo and heterodimerisation are equally likely, it was anticipated that about half the dimers formed would be C25/H25 heterodimers. Consequently if the heterodimer was truly active, that activity ought to be present in the resultant refolded solution. However, no activity was detected.

Figure 23 C25/H25 mutant heterodimer assay results.



0 M NaCl
Avg. Rel. Act.

pH	*Simple Mix*	*Urea Refold*	*Control*
4	-16.6	-16.4	7.6
5	-8.4	-13.9	8.9
6	-13.0	-8.2	5.0
7	-9.3	-2.6	-1.9
8	-13.1	-5.3	-4.3

1 M NaCl
Avg. Rel. Act.

pH	*Simple Mix*	*Urea Refold*	*Control*
4	-4.8	-6.6	16.3
5	-9.8	-11.7	16.9
6	-5.8	-12.7	12.4
7	-8.6	-7.9	-0.6
8	-9.2	-4.4	-1.5

See appendix tables 6-8 for data, pages 348-351.

The absence of detectable activity suggested either that any heterodimer formed was inactive, or possibly that the mutations prevented the intended heterodimer from forming (perhaps by disrupting the refolding process in some way). The formation of disulphide bonds between the Cys25 residues of the Cys mutant would have made homodimerisation of the Cys mutant more favourable than heterodimerisation. This undesirable situation was prevented by inclusion of 1 mM DTT in all buffers used.

Assay conditions

An optimum assay condition was not established since no activity was detected. If we assume that a heterodimer was formed, it was not active under the conditions favoured by native HIV-2 PR (fig. 23). It was anticipated that any heterodimer formed might be active in conditions preferred by papain – which includes a broad range of pH. However, a range of pH and salt conditions failed to elicit any signs of activity from mutant mixtures (fig. 23). Had activity been detected it would have been interesting to try to inhibit it using pepstatin and E64 (specific inhibitors of aspartic and cystein proteases generally).

The results suggest either that heterodimer formation did not occur, or that the heterodimer formed but was inactive. Given that a heterodimer was formed, the lack of activity observed could be interpreted in a number of ways. For example it may indicate either that an additional co-factor is required (an unlikely scenario since neither papain nor HIV-2 PR require co-factors), or that the heterodimer is not active at 37°C (also unlikely given that both HIV-2 PR and papain are active at 37°C). More realistically it may be that the heterodimer does not share the same substrate specificity as native HIV-2 PR (perhaps mutation of the active site has disrupted substrate binding), or possibly the correct spatial configuration of Histidine with respect to Cysteine has not been achieved and thus a functional papain-like active site is not present. The exact nature of any heterodimer formed could be determined from a crystal structure – assuming a crystal and X-ray data could be obtained. However, this would require the ability to separate homodimers from the heterodimer prior to crystal growth, and would also rely on the stability of the heterodimer during crystallisation. If either homodimer crystallised more readily than the heterodimer

then equilibrium in the crystal drop would shift in favour of homodimer crystal formation.

Another possibility is that the heterodimer was formed, and that it would have been active, but that activity was not seen due to oxidation of the thiol group during purification and subsequent analysis. However, DTT was used during preparation of the Cys mutant (except during the IMAC step) in order to minimise oxidation. Also oxidation is gradual so that any activity would have diminished with time, consequently some activity should have been detected even if oxidation was a problem since assays were carried out using freshly made protein.

The results can of course be explained if the intended His-Cys heterodimer was never formed. Two approaches to the formation problem were considered - first the mutants were simply mixed in equal quantity and assayed for activity; second they were denatured in urea, mixed in their denatured states then refolded together. The first approach relied on a natural dynamic equilibrium existing between the dimeric state and the monomeric state, the presence of such equilibrium has been exploited in the use of dominant negative mutations to form inactive HIV PR heterodimers (Babé, Pichuantes *et al.* 1991; McPhee, Good *et al.* 1996). Consequently a mixture of the mutant homodimers should eventually yield a mixed population of mutant homo and heterodimers due to monomer exchange. However, investigation of this equilibrium suggests that it conforms to a two-state model in which folded dimers are in equilibrium with unfolded monomers (Grant, Deckman *et al.* 1992). In such cases protein unfolding and dimer dissociation are concomitant processes, and the folded monomer cannot exist (Bowie and Sauer 1989). Given the inherent stability of the native dimer it seems probable that the homodimeric state would persist in preference to the unfolded monomer thus this exchange method would tend to be a particularly slow and inefficient means of generating the heterodimer. This concern, coupled with the experience of refolding native proteases with relative ease, lead to investigation of the refolding method for generation of heterodimer. A similar method has since been adopted and described by Rozzelle *et al.* (Rozzelle, Dauber *et al.* 2000) for the characterisation of HIV-1 PR based heterodimers.

An alternative solution to the problem of obtaining a heterodimer would have been to create a construct such that the two mutants were coexpressed from the same vector.

The two genes could be positioned one immediately following the other in frame, with a short intervening region encoding a flexible linker of perhaps 5 or 10 residues. A heterodimeric form of retroviral protease consisting of monomers from HIV-1 and HIV-2 respectively has been created using a similar linked system (Bagossi, Cheng *et al.* 1996) (Bagossi, Cheng *et al.* 1998). In this way a monomeric enzyme could be created having the papain active site but otherwise resembling HIV-2 PR. It is conceivable that (if active) this monomeric form could cleave its own linker region and so revert to a dimeric form. Indeed this could be deliberately engineered by selecting an appropriate sequence for the linker region and might incidently provide a useful indicator of activity, since an SDS gel could be used to distinguish an inactive product (which would run at approx. 22 kD plus linker size) from active linker free dimer (which would migrate at approx. 11 kD). A linked heterodimer co-expression system would probably represent the best method of ascertaining whether or not such a heterodimer is active. One drawback of these co-expression methods might be that (if active) the mutant heterodimer could prove to be just as cytotoxic as the native protease, and would not therefore be over-expressed to any great extent by the host cells.

In order to maintain the advantage of separate expression in terms of cytotoxicity, and yet confirm heterodimer formation in the absence of detectable activity, it would be necessary to differentially label or tag the monomers in some way. An option might be to His tag the Cys mutant whilst fusing the His mutant to Glutathione S-transferase (GST) for example, it may then be possible to isolate a heterodimer from solution by sequential purification steps. Fluorescent or perhaps radiolabelling of the respective mutants could also be used to detect heterodimer formation. Biacore experiments using a Ni-NTA chip with a His-tagged mutant versus a non His-tagged mutant could be used to detect and measure the extent of any heterodimerisation, and conceivably even to isolate small quantities of heterodimer by elution from the chip.

An alternative to the mutagenesis of HIV-2 PR would have been to mutate Papain. Changing Papain's catalytic Cys and His residues to Asp could achieve a similar goal, i.e. the conversion of the protease's mechanism to that of another class of protease by mutagenesis of the active site residues alone. Papain could be overexpressed in *E.coli* (or baculovirus) as ProPapain and subsequently activated, thus obviating problems of toxicity. Papain is monomeric, consequently there would be none of the problems

associated with generating a heterodimer. Without Cys as a catalytic residue the mutant Papain should not be subject to rapid loss of activity due to oxidation. Whether or not such mutagenesis would disrupt the folding of Papain is not clear, and this might prove to be the major obstacle to success. It is nevertheless an intriguing possibility.

SUMMARY

The mutants were successfully constructed, expressed, purified and characterised. As expected neither homodimeric mutant showed any proteolytic activity. Unfortunately no activity was detected on combination of the two different mutants either. The possibility that there was little or no dynamic equilibrium between the dimeric, monomeric, and heterodimeric state was considered; and the mutants were mixed while unfolded, and then refolded in the presence of substrate. No proteolytic activity was detected during these refolding studies. The possibility that the heterodimer was active under different buffer conditions (pH and salt) from the native enzyme was also investigated. Assays carried out using a range of possible conditions failed to detect any activity, despite the fact that papain is active over a broad range of conditions.

It was concluded that either the heterodimer is not easily formed, or that it has been formed but is not active. A more precise explaination would require further work, and potential actions were discussed.

The belief that retroviral proteases are toxic to cells in which they are expressed, and that this explains the poor yields observed has been questioned by the results of this chapter, and in particular by the difference in expression levels between the mutants. Results presented in the next chapter (Chapter 6) also suggest that there may be other influencing factors to consider. Eliminating toxicity was the primary aim in constructing an SIV_{AGM} PR expression system during attempts to express this protease for structural studies. The results were suprising, and lead to the investigation of codon useage as a primary cause of low SIV_{AGM} PR yields, and a possible cause of low retroviral PR yields generally.

ACKNOWLEDGEMENTS

Thanks to Dr. Paul Wan for recommending the QuickChange Site Directed Mutagenesis method, and for providing advise on its use.

Chapter Six: UDG-SIV_{AGM} Protease Fusion.

Development of an E.coli based protease expression system and preliminary crystallographic studies of the SIV_{AGM} Protease.

INTRODUCTION

The relationship between HIV and SIV.

The Simian Immunodeficiency Virus (SIV) is a species of lentivirus, closely related to HIV (Fig. 2, Chapter 1, page 26), which naturally infects wild monkeys (Kanki, Alroy *et al.* 1985; Ohta, Masuda *et al.* 1988). SIV is the closest known relative of HIV in terms of genetic, biological and antigenic similarities, making it suitable for the study of the pathogenesis of HIV infection, and for the development of vaccines (Kurth, Kraus *et al.* 1988). Subspecies of SIV have been isolated from a variety of primates. These include the mandrill, the rhesus macaque and the African green monkey – which each harbour distinct (i.e. primate species specific) forms of SIV, designated SIV_{MND}, SIV_{MAC} and SIV_{AGM} respectively.

The SIV_{MAC} (Chakrabarti, Guyader *et al.* 1987) and SIV_{AGM} [TYO-1] strain (Fukasawa, Miura *et al.* 1988) genomes have been sequenced as have those of HIV-1 [BRU] isolate (Wain-Hobson, Sonigo *et al.* 1985) and HIV-2 [ROD] isolate (Guyader, Emerman *et al.* 1987). They have also been compared with each other in terms of genomic organisation (Fukasawa, Miura *et al.* 1988). Comparison of these genomes has revealed that the SIV_{MAC} is more closely related to HIV-2 than to HIV-1 (Chakrabarti, Guyader *et al.* 1987; Franchini, Gurgo *et al.* 1987). In contrast it has also been shown that SIV_{AGM} is equidistantly related to both HIV-1 and HIV-2, and quite dissimilar to SIV_{MAC} (Fukasawa, Miura *et al.* 1988). It seems likely that these lentiviruses have diverged as their primate hosts evolved, and are now species specific (Fukasawa, Miura *et al.* 1988). In the case of HIV-1 and HIV-2 they have diverged from a common ancestor whilst maintained within separate isolated human populations.

SIV provides an animal model for HIV and AIDS.

Animal models are in use for the study of vaccines and drugs – studies on cats infected with FIV provide a small animal model (Schnolzer, Rackwitz *et al.* 1996),

whilst Macaque monkeys infected with SIV have been used for some time in vaccine trials (Almond, Jenkins *et al.* 1992) and are an ideal model for testing a variety of anti-retroviral drugs (Tomasselli, Bannow *et al.* 1992). The African Green Monkey provides an interesting model because SIV_{AGM} is present in a large proportion of wild African green monkeys, its natural host, yet the virus does not induce disease in them (Kraus *et al.* 1989, Allan 1991, Hartung *et al.* 1992, Hirsch and Johnson 1994, Norley 1996). Consequently studies of the African green monkey / SIV_{AGM} system may reveal the processes by which these monkeys remain unaffected by their parasite.

Other lentiviral proteases may provide an insight into the HIV drug resistance problem.

HIV has demonstrated an aptitude for the rapid development of resistance to drugs, and in particular to PR inhibitors (Ho, Neumann *et al.* 1995) (Gulnik, Suvorov *et al.* 1995) (Condra, Schleif *et al.* 1995). It has been demonstrated that greater than a third of the HIV-1 protease sequence is subject to drug-induced mutations (Schinazi, Larder *et al.* 1997). While the respective structures of retroviral PRs are very similar, there remain sufficient differences in terms of their interactions with substrates and inhibitors to provide possible insights into the nature of drug resistance in general (Wlodawer, Gustchina *et al.* 1995) (Gustchina, Kervinen *et al.* 1996) (Powell, Bur *et al.* 1996) (Slee, Laslo *et al.* 1995). This has provided a rationale for studies of the various retroviral PR structures. Of the SIV proteases, to date only the SIV_{MAC} PR has been solved (Wilderspin and Sugrue 1993; Wilderspin and Sugrue 1994), the work described in this chapter concerns the production of the SIV_{AGM} PR, for the purpose of future structural analysis by crystallography.

Problems associated with high-level expression of SIV_{AGM} protease.

The SIV_{AGM}PR gene was first cloned into several conventional expression vectors with a view to improving over expression and purification of the SIV_{AGM} protease for X-ray crystallographic analysis. SIV_{AGM}PR expression levels using these vector systems were extremely poor however, and the need for a new system of expression for the protease was recognised.

Development of a new protease expression system.

In pursuit of this aim a novel, *E.coli* based, retroviral protease expression system was developed and is described here. It was intended to be an improvement on existing systems in terms of yields, whilst allowing simplicity of detection and purification of

the protease. The system was developed primarily to allow the expression of SIV_{AGM} PR, although it is applicable in principle to other retroviral proteases. Ultimately it could be adapted to provide the basis for a low cost, high yield solution to the problem of producing retroviral proteases for industrial and research use. It comprises a fusion construct able to produce the protease in an insoluble, inactive, and thus non-cytotoxic, form. It allows purification by a single IMAC step, followed by refolding to yield active native SIV_{AGM} PR.

Following development of this system, codon usage was identified as the severely limiting factor in bacterial overexpression of this protease, and means by which this problem could be overcome are discussed.

The precise PR cleavage site within the SIV_{AGM} POL polyprotein at the N-terminal of the protease was identified and its significance is discussed.

MATERIALS AND METHODS

E.coli strains used

JM109 was used for all DNA preparation, BL21 (DE3) pLys S was used for all protein expression, except, as indicated in the text, where BL21-CodonPlus-RIL cells were used (supplied by Stratagene). The RIL strain cells carry extra copies of the argU, ileY, and ileW tRNA genes. The tRNAs encoded by these genes recognise AGA / AGG, AUA, and CUA codons respectively, which are the rarest codons in *E.coli.*

Source of the SIV_{AGM} PR gene

Professor Masanori Hayami kindly provided the SIV_{AGM} PR DNA template as a Bluescript phagemid. The phagemid incorporated an *XbaI-BamHI* subclone of a circularised SIV_{AGM} molecular clone, containing the entire *gag* open reading frame (ORF) and the first 1000 bases of *pol*. The coding sequence for a precursor form of the SIV_{AGM}PR (with extensions of 38 amino acids at the N-terminus and 26 amino acids at the C-terminus) was amplified from the phagemid by PCR, and inserted into the expression vector pMAL-cR1 to produce the construct designated pACS28 (A. Shearer, unpublished). All subsequent constructs incorporated PCR amplified gene products generated using pACS28 (fig. 1) as the template. The pACS28 fusion product is shown in figure 2.

Figure 1 Vector pACS28, encoding a fusion of the Maltose Binding Protein and an extended precursor of the SIV_{AGM} PR.

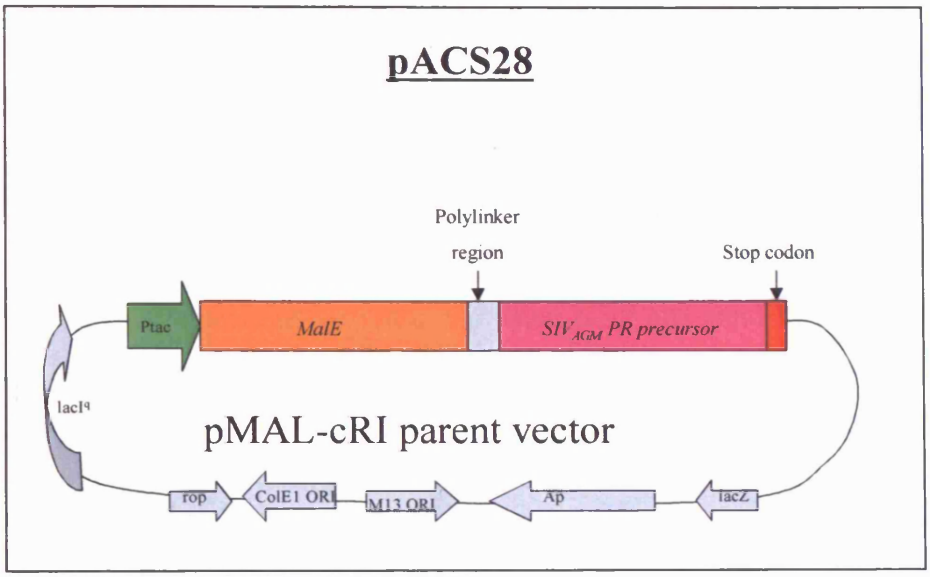
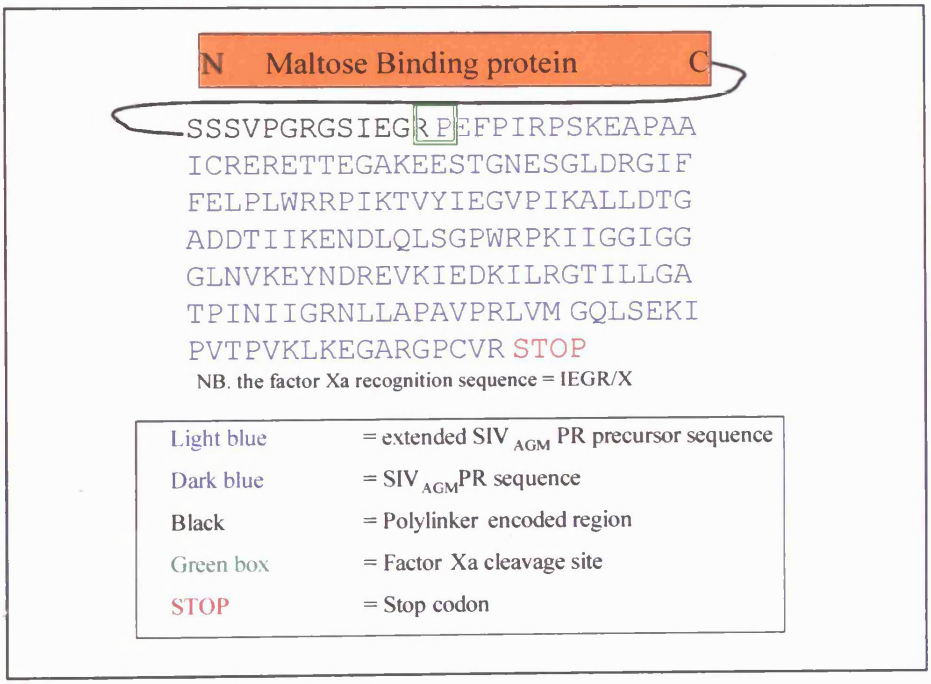


Figure 2 MBP-SIV_{AGM} PR-precursor fusion product encoded by pACS28.



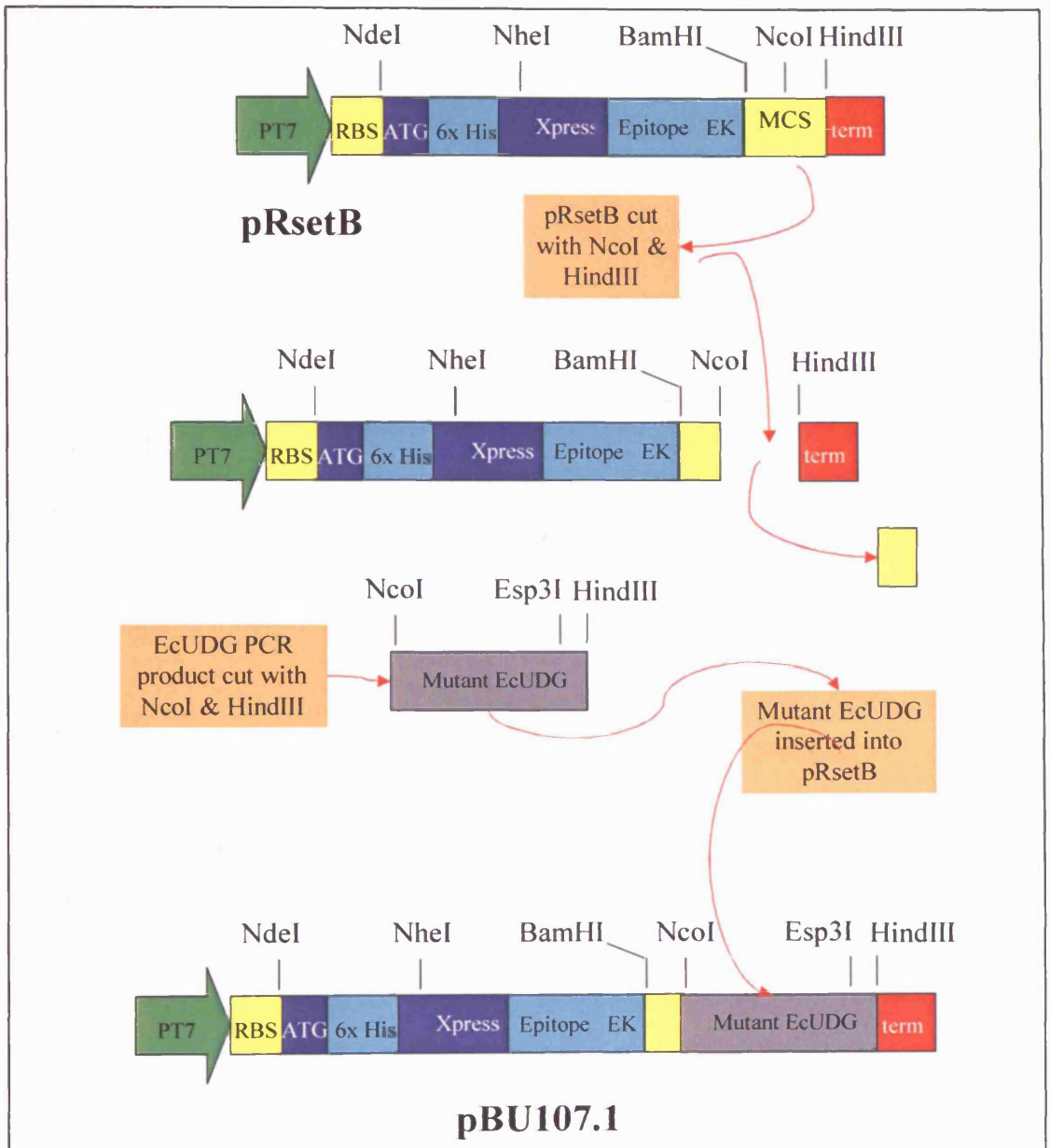
Source of the *E.coli* Uracil DNA Glycosylase (*Ec*UDG) mutant.

The pRsetB (Invitrogen) based plasmid pBU107.1 was constructed and provided by R.Savva (fig. 3). The use of a pRsetB based construct provided a strong T7 promoter and allowed single step purification of the expression product using an IMAC protocol. Further details of the pRset vector system, including a vector map and description of the polylinker region, are provided in the appendix (pages 340-341).

The *Ec*UDG gene was amplified by PCR using the primers to generate NcoI and HindIII restriction enzyme sites at the 5 and 3 prime ends respectively. The PCR product and pRsetB vector were each digested using NcoI and HindIII, and ligated to produce the construct pBU107.1. Expression studies revealed the *Ec*UDG gene product to be unexpectedly insoluble, leading to the assumption that the PCR product used to construct pBU107.1 was a mutant form of the *Ec*UDG gene (R.Savva personal communication).

Plasmid DNA was prepared by Qiagen midi-prep, as described in Chapter 2 (page 125). Plasmid pBU107.1 was used both for the expression of the mutant *Ec*UDG for control purposes, and as the basis for construction of the novel SIV_{AGM} PR fusion plasmid $pMIRA_6$, described in this chapter.

Figure 3 Construction of plasmid pBU107.1



PT7 = T7 promoter region
ATG = Start codon
MCS = Multiple Cloning Site
EK = Enterokinase cleavage site
RBS = Ribosome Binding Site
6 x His = PolyHistidine tag
term = Stop codon

The **Xpress Epitope** allows detection by antibody.

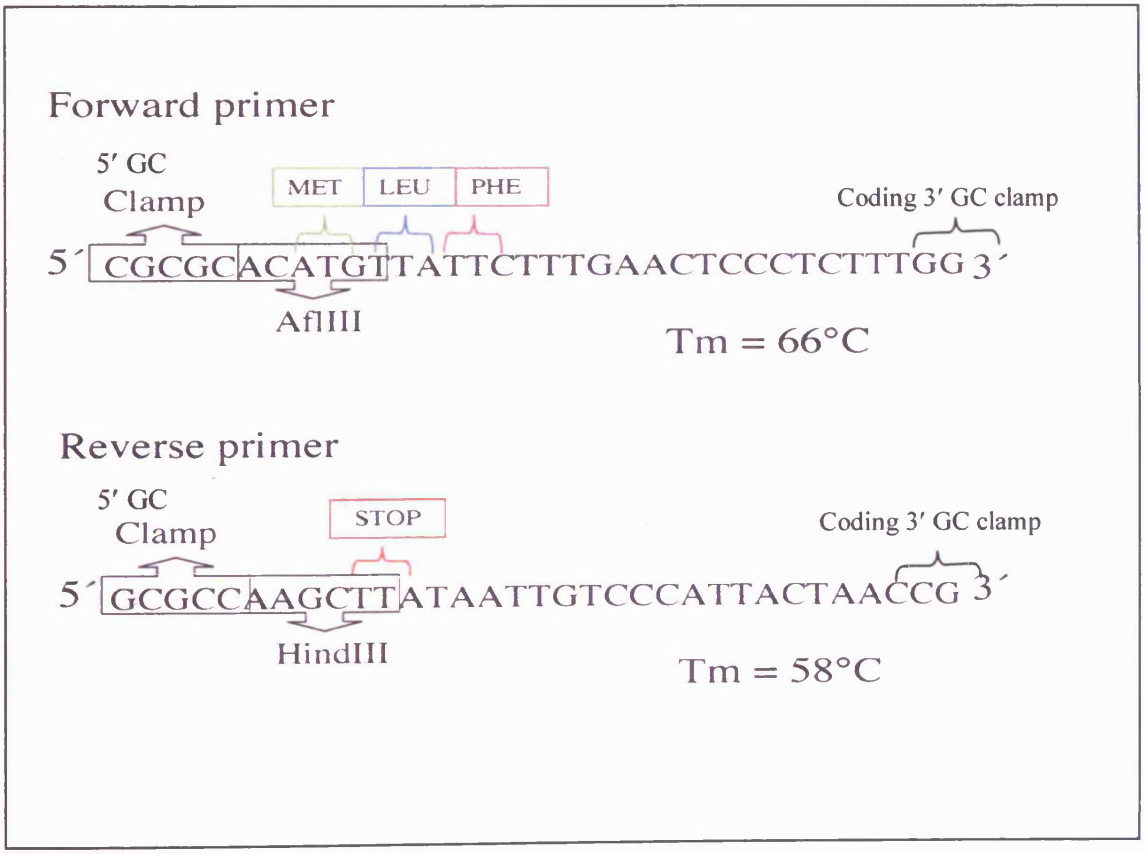
The above figure shows the construction of plasmid pBU107.1 by insertion of a mutant *E.coli* UDG gene into the multiple cloning site of the expression vector pRsetB via NcoI and HindIII restriction enzyme sites.

Construction of the EcUDG-SIV$_{AGM}$PR fusion vector, pMIRA$_6$.

Preparation of the SIV$_{AGM}$ PR gene by PCR

The SIV$_{AGM}$PR gene was amplified by PCR using the pACS28 construct as a template, and the primers illustrated in figure 4. The SIV gene contained an internal NcoI site, preventing the use of this enzyme for cloning purposes. AflIII generates overhanging ends that are compatible with those generated by NcoI, thus incorporation of an AflIII site into the forward PCR primer allowed insertion of the PCR product into any vector prepared with NcoI. Ligation of DNA ends prepared with AflIII and NcoI eliminated both parent sites in the process, but retained the original frame. The reverse primer incorporated a HindIII site and a stop codon. This strategy was used for insertion of the SIV$_{AGM}$PR gene into all of the various conventional expression constructs investigated, and whilst it could not be employed in construction of pMIRA$_6$, the same primers were nevertheless used for amplification of the initial SIV gene product.

Figure 4 SIV$_{AGM}$ PR PCR primers, as used to construct pMIRA$_6$

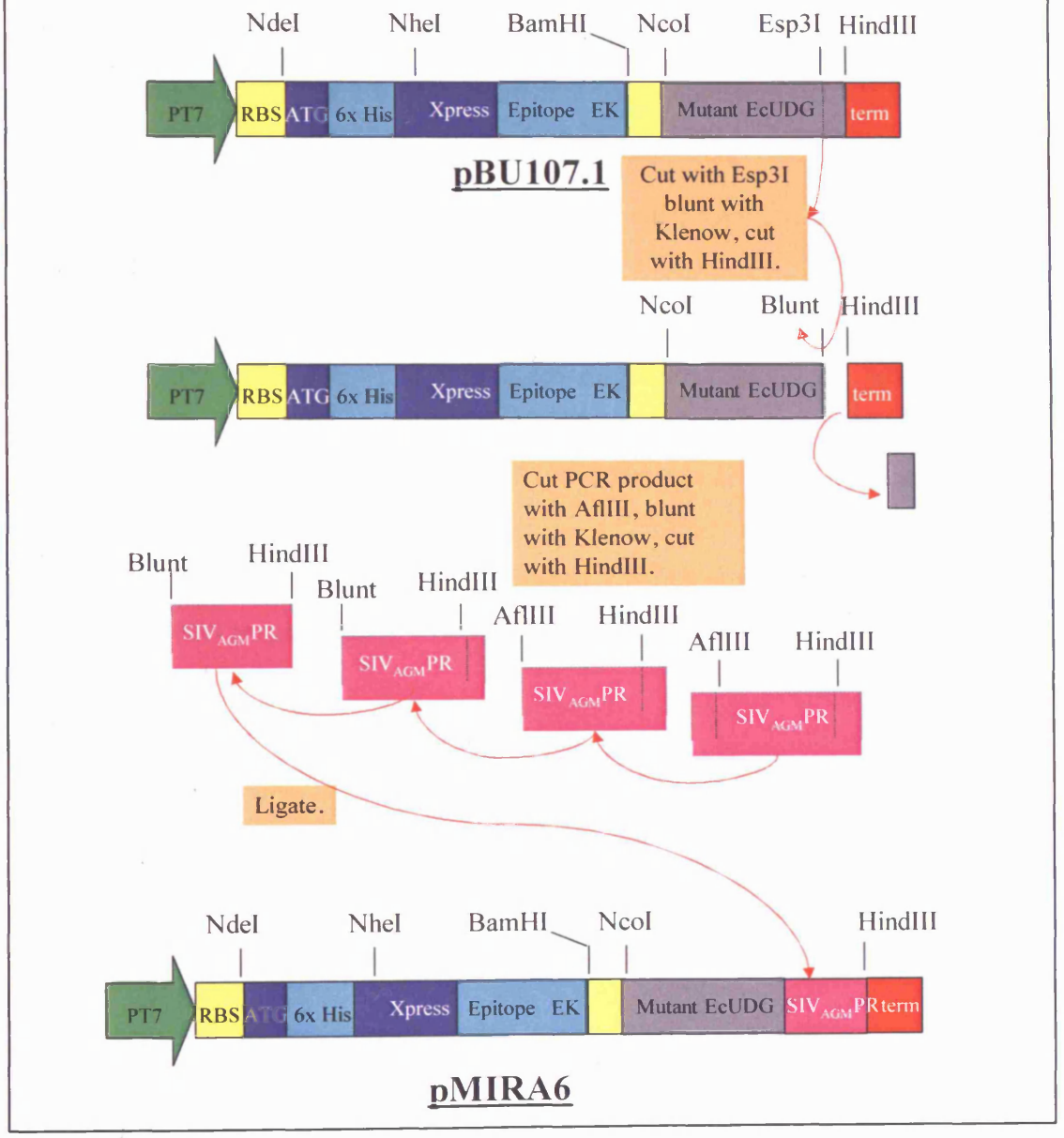


The cloning strategy

The *Ec*UDG gene contained an Esp3I site at a position 70 bp from the 3′ end. This was the site selected for insertion of the SIV_{AGM} PR PCR product into pBU107.1. The cloning strategy is illustrated in figure 5. The plasmid pBU107.1 was first digested with Esp3I (from Fermentas) and exposed to Klenow in order to generate blunt ends. Next the distal 70 bp of the 800 bp UDG gene was removed by digestion with HindIII. Meanwhile the SIV_{AGM}PR gene PCR product was digested at the 5′ end with AflIII, then exposed to klenow generating a 5′ blunt end. The SIV gene was subsequently digested with HindIII to provide a 3′ overhang. Following their preparation both plasmid and insert were purified by band extraction from an agarose gel. The 350 bp SIV_{AGM} PR gene was then inserted by ligation into the UDG construct, such that the blunt 5′ end of the SIV gene ajoined the blunt 3′ end of the shortened UDG gene. The SIV_{AGM} PR gene was thus positioned in frame with the 3′ end of the truncated UDG gene so that uninterupted read-through would occur during translation. The procedures for digestion, use of klenow, and ligation are described in Chapter 2 of this thesis (see DNA manipulations, pages 134-137).

Figure 5 The construction of pMIRA6.

The SIV_{AGM} PR PCR product and pBU107.1 were prepared as described in the text. The SIV gene was then inserted into the plasmid downstream and in frame with the recently truncated EcUDG gene.



Isolation of clones containing the desired UDG-SIV fusion construct.

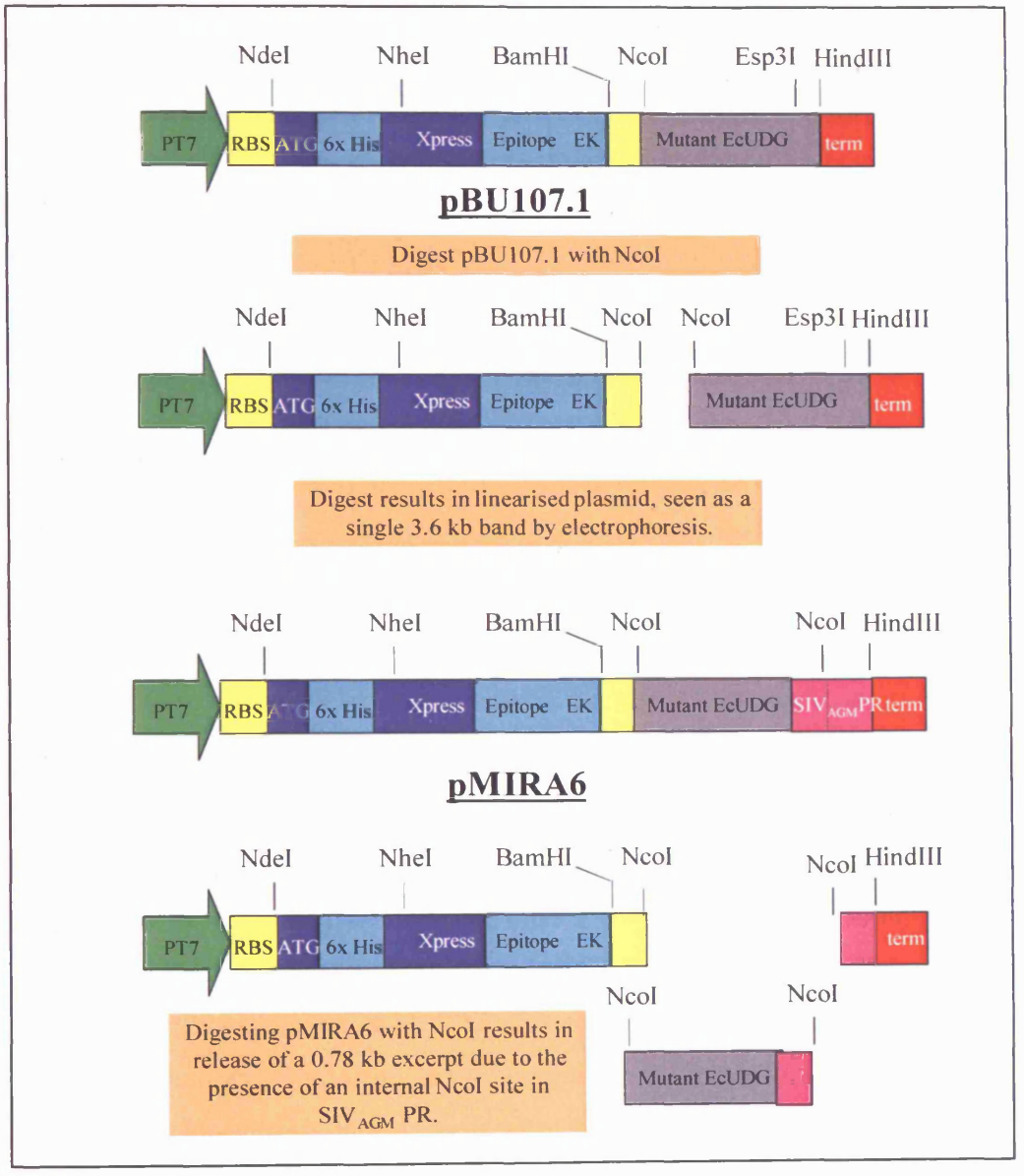
Transformation of *E.coli* JM109 cells with DNA of the final construct resulted in numerous colonies which were screened for presence of the correct insert by the following methods:-

1. Restriction digest with NcoI, to yield two bands at 3.1 kb and 0.78 kb where the SIV-PR gene was present, or a single 3.6 kb band in the case of uncut or religated vector (illustrated in fig. 6 and results shown in fig. 12).
2. PCR amplification of the SIV_{AGM} PR gene, if present[13] (results, fig. 13).
3. Direct sequencing of the DNA construct using a Pharmacia T7 sequencing kit and the SIV-PR 3' PCR primer (results, fig. 14).

Double restriction digest with both NdeI and HindIII was not used as a screen since the correct construct would yield two bands of 2.8 kb and 1.0 kb, whilst uncut or religated pBU107.1 would yield two bands of 2.8 kb and 0.8 kb. These two results may have proved difficult to distinguish. Competent cell production, transformations, restriction digests, PCR reactions, agarose gel electrophoresis and DNA sequencing procedures were all carried out as described in Chapter 2 (see pages 138, 141, 134, 128, 134, and 130 respectively).

[13]Using primers described in figure 4 of this chapter.

Figure 6 NcoI digests indentified the correct pMIRA construct.



Small scale expression screens.

These were carried out as described in Chapter 2, using the BL21 (DE3) strain and BL21 (DE3) Codon-Plus-RIL strain.

Expression and purification of the mutant *Ec*UDG control protein.

The mutant *Ec*UDG was overexpressed in *E.coli* strain BL21 (DE3), containing the pBU107.1 plasmid. Cultures (1 litre) were prepared, grown (at 30°C), induced, and harvested as described in Chapter 2 (page 147). Cells were resuspended in 50 mM Tris pH 8.0, 100 mM NaCl and stored at -20°C pending use. Cells were thawed on ice to lyse (exploiting the activity of lysozyme encoded by the pLys S plasmid) and a brief sonication step reduced viscosity and ensured complete lysis (Chapter 2, page 151). The lysate was clarified by ultracentrifugation at an RCF of 141000 g for 1 hour at 4°C. The clarified supernatent was discarded and the pellet fraction was briefly washed by resuspension in 50 mM Tris pH 8.0, 100 mM NaCl to remove any residual soluble protein, and re-pelleted by ultracentrifugation as before. Insoluble proteins in the pellet were dissolved by rotary incubation at 4°C overnight in 50 mM Tris pH 8, 100 mM NaCl, containing 8 M deionised urea. Stubborn lumps of undissolved pellet were dispersed by brief sonication. Undissolved material was removed by ultracentrifugation at an RCF of 141000 g for 1 hour at 4°C, and the clarified supernatent was retained for loading onto a 50 ml histidine binding IMAC column pre-equilibrated in the same Tris / urea buffer. Purification by IMAC was carried out (see Chapter 2, page 159), with 8 M deionised urea present in all buffers. Urea was the chosen chaotrope in preference to guanidium HCl on grounds of reduced cost. The flow through, wash and elution peaks were collected for analysis by PAGE.

Analysis of the mutant EcUDG's susceptibility to cleavage by PR.

An aliquot (10 µl) of the pure mutant *Ec*UDG was micro-dialysed (chapter 2, page 158) against each of the HIV-1 and HIV-2 PR activity buffers at 4°C overnight. Following dialysis they were transferred to eppendorfs, and 1 µl of the appropriate HIV PR was added to each aliquot. An additional aliquot was also dialysed against HIV-1 PR buffer, but no HIV enzyme was subsequently added to it. This tube acted as a control. The tubes were then incubated at 37°C overnight. The *Ec*UDG was subsequently assessed for signs of degradation by Western blotting (Chapter 2, page 153) using primary antibodies raised against the N-terminal 6x histidine tag present on the *Ec*UDG (fig. 8).

Expression and purification of the mutant *Ec*UDG-SIV$_{AGM}$ PR fusion protein.
Large-scale (8 L) shake flask cultures were grown at 30°C, induced and harvested as for the mutant *Ec*UDG protein (above). Cell lysis, preparation of insoluble protein from the pellet fraction, and subsequent IMAC purification of the fusion protein were also carried out as described above for the mutant *Ec*UDG protein (results fig. 19).

Concentration of the mutant *Ec*UDG-SIV$_{AGM}$ PR fusion after IMAC.
The fusion protein was concentrated by removal of chaotrope through a "cold refold" dialysis step. This concentration procedure exploited the fact that an aspartic protease would be inactive at pH 8 and 4°C, and thus unable to release itself from the fusion protein under these conditions even if it refolded. The precipitated fusion protein was then harvested and redissolved in a reduced volume of the original chaotropic buffer.

The eluent from the IMAC step (containing the mutant *Ec*UDG-SIV$_{AGM}$ PR fusion in 50 mM Tris pH 8 containing 8 M urea, 100 mM NaCl, and 300 mM immidazole) was dialysed against 50 mM Tris pH 8 containing 100 mM NaCl, 1 mM EDTA, at 4°C overnight. The insoluble fusion protein dropped out of solution as the urea was removed and was then collected by ultracentrifugation at an RCF of 141000 g for 1 hour at 4°C. The pellet was redissolved in a small volume (approx. 5 ml) of 50 mM Tris pH 8 containing 8 M urea.

Concentrating the fusion in this way prior to warm refolding maximised the efficiency of subsequent proteolytic cleavage. An additional benefit of having two refolding steps (cold and warm) was the removal (during the cold step) of any insoluble proteins that may otherwise have refolded in concert with the protease and contaminated the final preparation of soluble protease.

Autocatalytic cleavage and release of the SIV$_{AGM}$ PR from within the fusion, by warm refolding.
The concentrated fusion protein solution in urea, resulting from the steps above, was dialysed against 50 mM Sodium acetate pH 4.5, 100 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM EDTA at 37°C for 2 hours. As the urea concentration subsided a precipitate formed and was removed by centrifugation for analysis by SDS-PAGE. The supernatent solution following dialysis contained the refolded, active protease. The pure SIV$_{AGM}$ PR was concentrated by Centricon then stored at -20°C pending further characterisation. Release of the protease fragment from the fusion was visualised by SDS-PAGE (results fig. 19).

Characterisation of the active, refolded, SIV_{AGM} protease

PAGE analysis.

The concentrated SIV_{AGM} PR was analysed by PAGE (results fig. 19), and visualised by coomassie staining.

Confirmation of aspartic protease activity by inhibition of autocatalysis.

Inhibition of protease release from the fusion product was investigated by inclusion of 10 μM pepstatin within the dialysis buffer during the warm refolding step (results fig. 19).

Immunological characterisation of SIV_{AGM} PR.

Dot blots (Chapter 2, page 153) were carried out using primary antibodies raised against HIV-1 PR and HIV-2 PR.

N-terminal analysis and Mass Spectroscopy.

Aliquots of the pure protein were subjected to N-terminal analyses and Mass Spectrometry, by the Chemistry department, University College London.

Crystallographic analysis of SIV_{AGM} PR

Hanging drops (2 μl), using the conditions which had been successful in the crystallisations of HIV-1 PR (McKeever, Navia *et al.* 1989, and Chapter 3, pages 189 and 197) and HIV-2 PR (see Chapter 4, pages 208, and 222-228) were set up as described in chapter 2 (page 170), using the pure active SIV_{AGM} PR at a concentration of 6 mg/ml, in a ratio of 1:1 with the crystallisation buffer, and incubated at 14°C. The drops were observed regularly using a light microscope, over a period of 1 month.

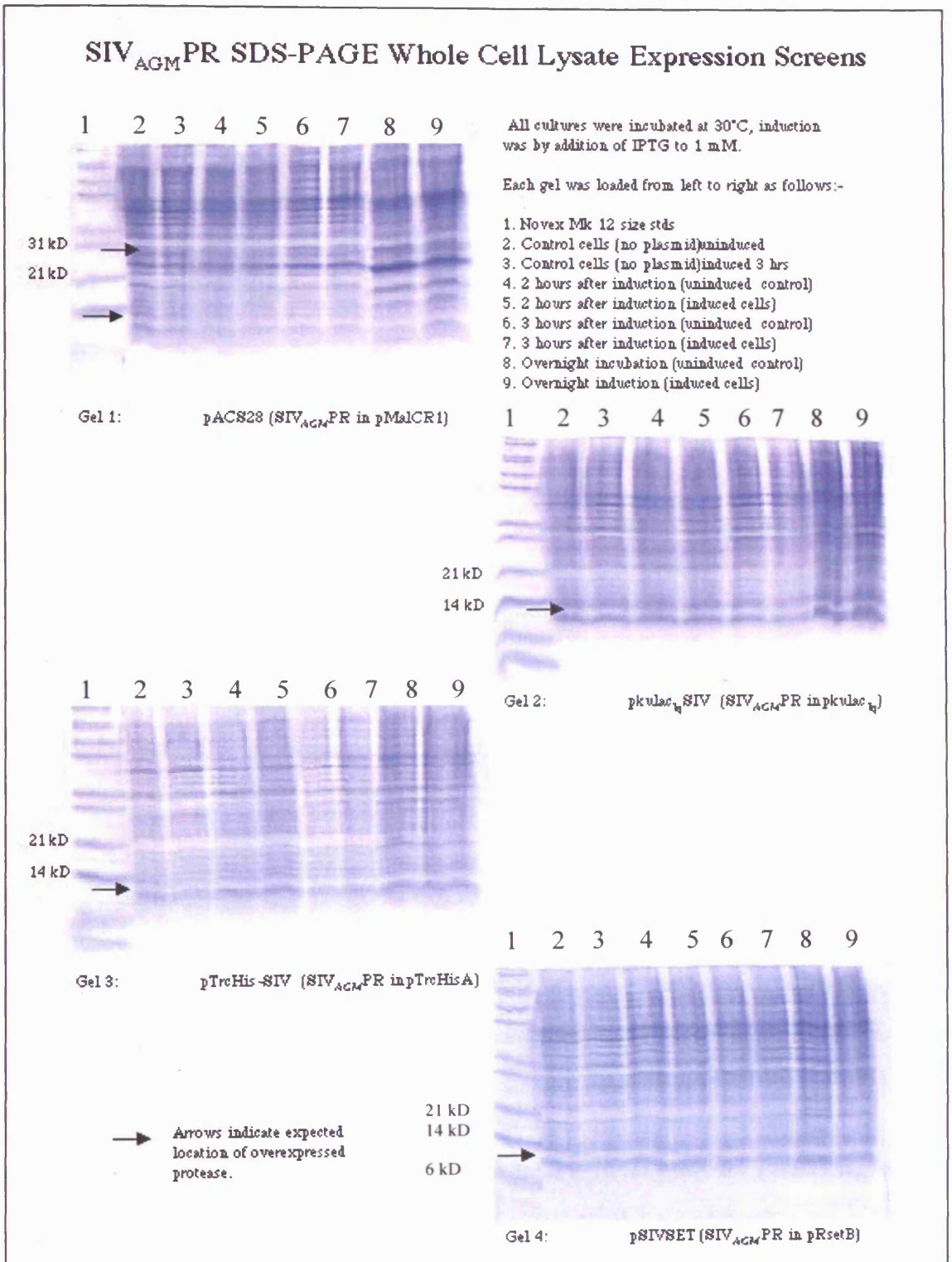
RESULTS AND DISCUSSION

The pMAL-cR1 vector (from NEB) provides a strong *tac* promoter to drive the expression of a heterologous protein as a C-terminal fusion with the *E.coli* Maltose Binding Protein (MBP). The construct pACS28 (A.Shearer) was intended to provide an SIV_{AGM} PR expression system, and utilised pMAL-cR1 as the recipient vector. Following overexpression of the MBP-SIV_{AGM} PR fusion in *E.coli*, active site titrations using pepstatin-A had been carried out to determine the yield of SIV_{AGM} PR obtained and it was estimated that the total protease yield was 0.75 mg per litre of cell culture, with 91% of protease activity recovered from the insoluble fraction and only 9% recovered from the soluble fraction (A.Shearer, personal communication). Furthermore, only 10% of total protease activity could be isolated by means of binding to and elution from amylose resin via the MBP moiety (A.Shearer, personal communication) indicating that 90% of the protease had achieved release from the fusion protein *in vivo*. Since the use of an MBP fusion did not provide a significant advantage in terms of protease expression levels, solubility, or ease of purification it was decided to investigate alternative expression systems.

Expression of SIV_{AGM} PR from conventional bacterial systems.

Production of SIV_{AGM} protease using a variety of commercially available protein expression systems was investigated, but in all cases protease expression levels were exceptionally poor. None of the commercial systems were any improvement on the existing HIV expression system (described in Chapters 3 and 4) either in terms of yield, or in terms of purification procedures. These investigations are not reported in detail here, however, an SDS-PAGE analysis of whole cell expression screens representative of these investigations is shown to illustrate the point (fig. 7). Such poor expression makes purification extremely difficult and in the case of SIV_{AGM} this was exacerbated by its ability to cleave itself from conventional fusion or tagged systems. The absence of an SIV_{AGM} PR specific antibody for Western blotting and the fact that the SPA system was no longer available to our laboratory, made detection of the protease during stages of expression and purification more difficult too.

Figure 7 Conventional bacterial expression of SIV_{AGM} PR gives low yields.



Overexpressed protein bands can be detected by SDS-PAGE analysis of whole bacterial cells using Coomassie stain, and this is a good indication of protein yield. In the examples above protein expression was induced in various conventional systems incorporating the SIV_{AGM} PR gene. Subsequent SDS-PAGE of whole induced bacterial cells failed to reveal any overexpression of the protease visible by Coomassie staining, indicating that yields were extremely low.

The toxicity problem

The expression of HIV-1 protease in living bacterial cells results in their loss of viability (Korant and Rizzo 1991). The degradation of cellular proteins in *E. coli* expressing large amounts of active HIV protease has also been observed (Korant and Rizzo 1991). The expression of the HIV-1 protease was found to be toxic to *E. coli* strain BL21(DE3) (Baum, Bebernitz *et al.* 1990). A strong correlation between the presence of functional protease and toxicity has been both demonstrated and exploited as a selection procedure for the identification of protease mutations (Baum, Bebernitz *et al.* 1990).

The possibility that the SIV_{AGM} protease may be toxic to the bacterial cells in which it is expressed due to cleavage of host cell proteins by the protease was therefore considered as an explanation for the low yields of protease expressed using conventional vector systems. Innocuous proteins generally allow high cell densities to be achieved during host cell culture, and are overexpressed with yields of 20 – 60% of total cell protein (i.e. 8 – 24 mg protein per litre of cell culture) (Huang, Newton *et al.* 1987; Craig, Yuan *et al.* 1991). However, cytotoxic proteins often limit the cell density that can be achieved in culture, and are poorly expressed with yields of 0.75 – 2.5% of total host cell protein (i.e. 0.3 – 1 mg protein per litre of cell culture) (Danley, Geoghegan *et al.* 1989; Menendez-Arias, Young *et al.* 1992).

Protein expression can be assessed in comparative terms by means of SDS-PAGE analysis. Commercial size standard markers (e.g. novex mark 12) are intended to be clearly visible following coomassie staining and are typically formulated such that they approximate to 0.5 μg of protein per band, given that the recommended volume per lane is loaded. *E.coli* produce approximately 40 mg total cellular protein per litre of culture, when grown to an OD_{600} of 1.0. If cells from 1 ml of culture (at an OD_{600} of 1.0) are harvested, then resuspended in 250 μl, and 10 μl of the resultant suspension is loaded per lane onto a gel, the total cellular protein in each lane would be approximately 16 μg. This means that a band of 0.5 μg (i.e. one that is clearly visible, and roughly comparable to the size standard bands in quantity) represents just over 3.1 % of total cell protein – or about 1.25 mg per litre of culture. It is generally considered that a yield of 1 mg per litre of culture is the practical minimum desirable for the commencement of structural studies. At yields lower than this very high volumes of cell culture would be required to produce sufficient protein for the

screening, optimisation, and refinement processes involved in solving a crystallographic structure. Thus if a clear band of the desired protein is not visible by coomassie staining (given the gel loading parameters described here), then there is unlikely to be sufficient yield of that protein to commence structural studies. For the purposes of appraising the relative effectiveness of bacterial expression systems, analysis by means of SDS-PAGE and coomassie staining is a fast, simple, and convenient technique and has been used here to compare the expression of proteases using a variety of systems.

The effect of protease activity on yield was witnessed during expression of HIV-2 PR active site mutants described in the previous chapter. The mutation of a single amino acid (Asp 25) to knock out protease activity resulted in an immediate improvement in protease yield (Chapter 5). Similar results had been reported elsewhere (Baum, Bebernitz *et al.* 1990). The problem remained, how to achieve a high yield whilst also producing an active native protease as the final product?

HIV-1 PR was expressed in an insoluble form as inclusion bodies, whereas HIV-2 PR was expressed in solution from the same system (work presented in previous Chapters). Yields of HIV-1 PR were generally observed to be moderately better than HIV-2 PR, and this was attributed to the inactivity of HIV-1 PR in inclusion bodies. A protease expressed as an insoluble, misfolded aggregate is prevented from dimerising and thus from becoming active. It had been established that these proteases could be easily refolded from chaotropic solutions. Insolubly expressed HIV-1 PR can be released with 8 M urea from washed cellular inclusion bodies, and after refolding such preparations cleave HIV-1 PR substrate with a specific activity comparable to column-purified HIV-PR (Cheng, McGowan *et al.* 1990). This was a procedure exploited in the production of both HIV-1 PR and HIV-2 PR (Chapters 3 and 4), and during the attempts to create an HIV-2 PR heterodimeric mutant (Chapter 5). Protease activity was detectable immediately upon refolding. A solution of unfolded protease in guanidine or urea can be dispensed into a larger volume of an activity buffer and protease activity is detected at once (this thesis). It was therefore considered that if the protease could be forced into an insoluble state by fusing it to a naturally insoluble protein it may still be possible to achieve release of the protease from the fusion by autocatalysis after refolding it from a chaotropic solution under favourable conditions for activity. The protease would be rendered inactive, and consequently non-toxic,

until deliberately refolded following purification. Selection of a fusion partner protein that was highly expressed in *E.coli* was also calculated to improve the yield of fusion and thus protease expressed.

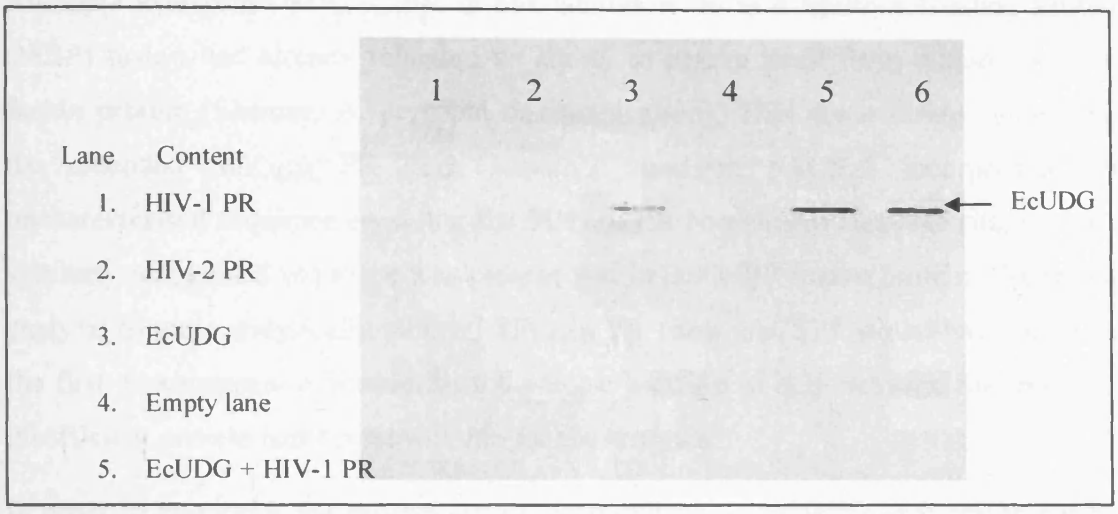
The selection of a mutant *E.coli* Uracil DNA Glycosylase as the fusion protein

The characteristics required of the fusion protein were that it should not be a substrate for the retroviral protease, it should be highly insoluble, not likely to refold, and it must be expressed at high levels in *E.coli*. To achieve high level insoluble overexpression in *E.coli* the natural choice is a highly abundant insoluble *E.coli* protein.

Uracil DNA glycosylases are ubiquitous enzymes able to recognise and remove misincorporated uracils from single and double stranded DNA, thus initiating a base excision repair pathway. These enzymes were the subject of other projects being pursued within our laboratory concurrent with this thesis, and have been extensively reviewed elsewhere (Parikh, Putnam *et al.* 2000; Pearl 2000; Scharer and Jiricny 2001).

Plasmids encoding the *E.coli* Uracil DNA Glycosylase (*Ec*UDG) for overexpression in *E.coli* were available in our laboratory. The *Ec*UDG derived from these systems was highly expressed, however, it was also highly soluble. By chance a mutant *Ec*UDG gene was isolated during the construction of other plasmids, and expressed an insoluble form of the protein (Savva, R. personal communication). The mutant *Ec*UDG was expressed at exceptionally high levels in *E.coli* but remained completely insoluble and resisted all attempts to refold it (Savva, R. personal communication). It was subsequently demonstrated to be totally resistant to cleavage by the proteases from HIV-1 and HIV-2 (fig. 8). This is perhaps not suprising since it has recently been shown that HIV-1 is able to package human UDG into viral particles by means of interactions between the UDG and the HIV-1 IN protein (Willetts, Rey *et al.* 1999).

Figure 8 Incubation of mutant EcUDG with HIV proteases.



Since Lentiviruses from primate species do not contain UDG encoding sequences in their genomes it is possible that incorporation of the host enzyme may be a ploy to regulate uracil misincorporation into the viral genome. Since the virus packages an RNA genome presumably activity by captured human UDG would be utilised in the cytoplasm of a subsequent host following reverse transcription of the RNA genome into DNA prior to integration. Indeed, this may explain why the interaction is specifically with IN – the enzyme responsible for integrating viral DNA into the host cell's genome. The virus would need to package UDG in this scenario, since the enzyme would be absent from the host cell cytoplasm (it is present only in the nucleus). Furthermore it is likely that since none of the primate lentiviruses encode their own UDG that this system of enzyme capture pervades them all – including SIV. Given that human UDG is packaged within HIV-1 viral particles and is required to ensure the integrity of the viral genome, a selection pressure exerted on the virus in favour of a protease that does not cleave human UDG is likely. It may be that the similarities between human UDG and *E.coli* UDG are sufficient to ensure that the bacterial enzyme is also resistant to the retroviral proteases. Whatever the reason behind *Ec*UDG resistance to PR, this characteristic, coupled with the others discussed, made it an ideal candidate for the role of a fusion protein in the development of a new SIV_{AGM} PR expression system.

Design of the fusion junction to allow protease release

Attempts to express SIV_{AGM} PR in our laboratory from a Maltose Binding Protein (MBP) fusion had already revealed its ability to cleave itself from within a soluble fusion protein (Shearer, A. personal communication). This demonstrated either that the extended SIV_{AGM} PR gene sequence used in pACS28 incorporated the uncharacterised sequence encoding the SIV_{AGM} PR N-terminal cleavage site, or that a similarly recognised sequence was present within the MBP fusion protein. N-terminal analysis of autocatalytically cleaved SIV_{AGM} PR from pACS28 would have provided the first experimental evidence for the precise location of this cleavage site, however insufficient protein had been available for the analysis.

In order to maximise the probability of autocatalysis on refolding, the fusion needed to incorporate the region most likely to contain protease cleavage site 5 (see fig. 10 Chapter 1). The precise locations of SIV_{AGM} PR cleavage sites within the SIV_{AGM} polyproteins were not known, nor was any information concerning SIV_{AGM} PR substrate specificity available from the literature. However, this information was available for HIV-1 protease, and had been predicted for HIV-2 PR and SIV_{Mac} PR (fig 11, Chapter 1). Consequently inferences could be made concerning the likely sites cleaved by SIV_{AGM} PR. Initially a basic sequence alignment between the HIV and SIV proteases was carried out, and the cleavage sites already known were used as an indication of the possible positions of those that were not. For the purposes of fusion protein construction the most important site was number 5 (fig 11, Chapter 1, page 60), releasing the N-terminal end of the protease from within the SIV_{AGM} POL polyprotein. Using the alignment in combination with the information available with respect to substrate specificity for HIV-1 PR (fig 12, Chapter 1, page 60) the likely position of cleavage site 5 in SIV_{AGM} was predicted to be between two phenylalanines (fig. 9). A structure-based alignment of three retroviral proteases (HIV-1, HIV-2, and SIV_{MAC}) with the amino acid sequence of the SIV_{AGM} PR (as expressed by the fusion construct) has subsequently been generated (fig. 10) using the bioinformatics programs Comparer, Clustal W and JOY (colours as per (Taylor 1997)) and concurs with the prediction originally made.

Figure 9 Preliminary sequence alignment and prediction of SIV_{AGM} PR cleavage site numbers 5 and 6.

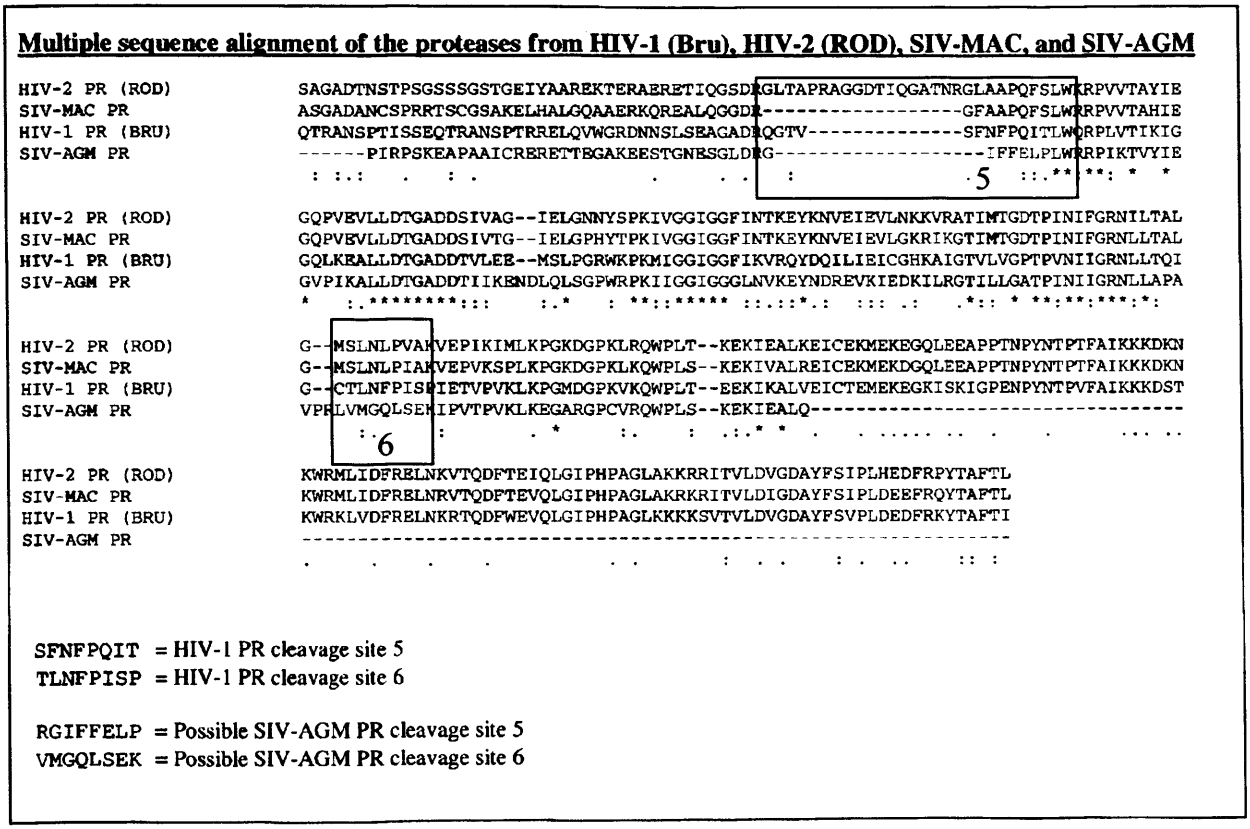


Figure 10 Structure based alignment supports initial SIV_{AGM} PR cleavage site predictions.

The program "Comparer" was used to align the structures of the proteases from HIV-1, HIV-2, and SIV_{MAC} using their respective PDB files. The result was then used to produce a structure based alignment of these proteases with the sequence from SIV_{AGM} PR using "ClustalW". The results were formatted for presentation using "JOY". The SIV_{AGM} PR residues are presented in the Taylor colours (Taylor 1997).

Key:

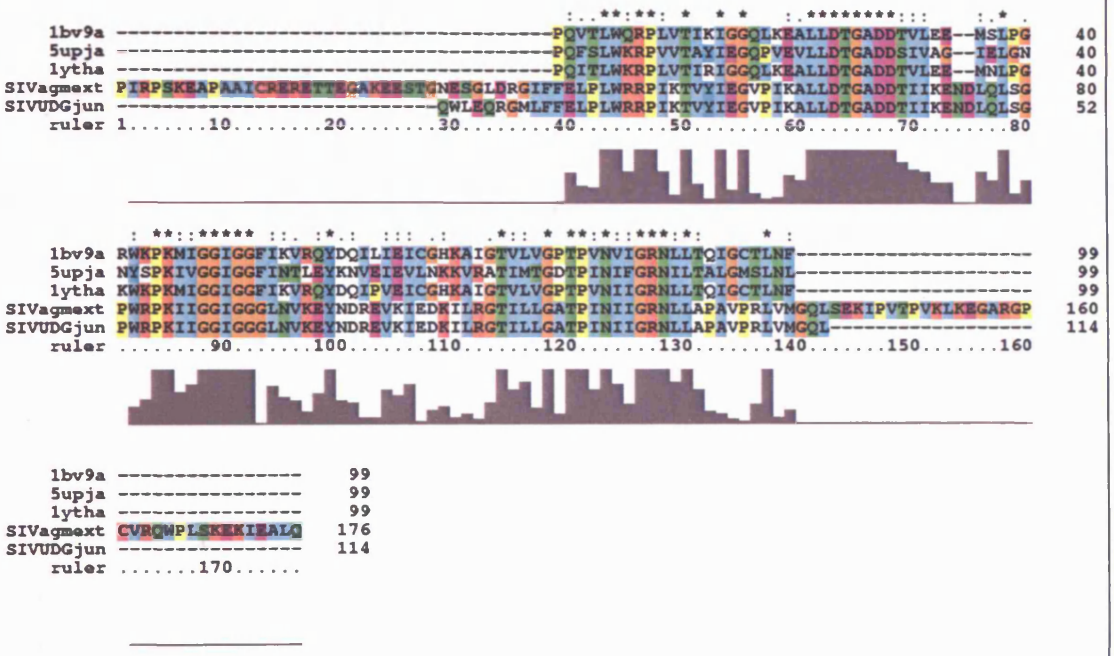
alpha helix	red
beta sheet	blue
solvent accessible	lower case
solvent inaccessible	UPPER CASE
hydrogen bond to main-chain	**bold**
hydrogen bond to main-chain carbonyl	underline
positive phi torsion angle	*italic*
Consensus secondary structure	a = alpha helix b= beta sheet

Figure 11 **Clustal W structure-based alignment of retroviral proteases with the SIV_{AGM} PR precursor and the *Ec*UDG-SIV_{AGM}PR fusion junction.**

1bv9a = HIV1 PR 5upja = HIV-2 PR 1ytha = SIV_{MAC} PR

SIVagmext = SIV_{AGM} PR precursor as cloned into pACS28

SIVUDGjun = SIV_{AGM} PR and junction with the C-terminal region of UDG



Summary of N-terminal protease cleavage sites.	P P′ 4321:1234
HIV-1 PR	SFNF:PQIT
HIV-2 PR	GLAA:PQFS
SIV-MAC PR	GFAA:PQFS
SIV-AGM PR	RGIF:FELP
EcUDG/SIV-AGM PR	GMLF:FELP

If the alignments based on the known sites for HIV-1 had correctly identified cleavage sites 5 and 6 for HIV-2, SIV_{MAC} and SIV_{AGM}, then the result could be used to assess the new sequence at the junction created between *Ec*UDG and SIV_{AGM} PR in the fusion protein. The probable SIV_{AGM} PR site 5 recognition sequence was RGIF:FELP, and this was compared with the UDG:SIV fusion sequence of GMLF:FELP.

Position P1

At P1 in both cases there was a phenylalanine, consequently no effect on cleavage efficiency was envisaged. Furthermore HIV-1 PR also had a phenylalanine at this position.

Position P2

At P2 there was a change from isoleucine to leucine. These residues are isomeric and they are similarly sized, non-polar, uncharged aliphatic residues. This change would probably not prove problematic. In terms of HIV-1 PR consensus for this position the most important parameter was that P2 residues were never charged, and both the isoleucine and leucine fulfilled this criteria.

Position P3

P3 represented the most disparate change, from glycine to methionine. There was an obvious size difference, with glycine being much smaller than methionine. Otherwise, both were non-polar, uncharged, non-aromatic residues. The HIV-1 PR consensus for P3 was for a small non-polar residue, which fit glycine but perhaps excluded methionine by virtue of size. However, HIV-1 PR actually had a phenylalanine at P3 of site 5, and one was also predicted by alignment for SIV_{MAC} PR, while HIV-2 PR was predicted a leucine. Leucine is of roughly comparable size to methionine, whilst Phenylalanine is slightly larger. Consequently it was considered that the methionine might be tolerated in this position.

Position P4

The final change was at P4, where arginine was exchanged for glycine. There was no conflict in terms of exclusion by size, since glycine is smaller than arginine. The general consensus for HIV-1 at the P4 position was for a very small residue, usually serine or threonine, consequently glycine should be tolerated here. In addition a glycine was predicted at residue P4, site 5, for HIV-2 PR and SIV_{MAC} PR sequences, suggesting that this may be an acceptable change. Finally a charged residue was never

seen at P4 in any of the HIV-1 sites, and since arginine is a charged residue the exchange at this position may ironically have made the SIV_{AGM} PR cleavage more efficient.

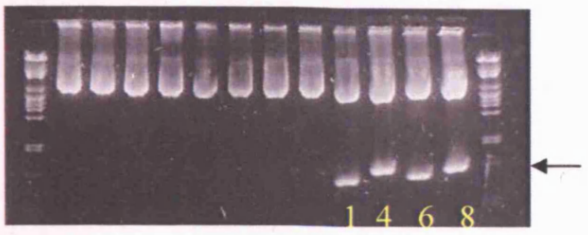
These analyses suggested that insertion of the protease gene at the Esp3I site of *Ec*UDG would, on translation, create a UDG-SIV junction compatible with the characteristics of the anticipated natural N-terminal cleavage site of the SIV_{AGM} PR and similar to the known sequences for retroviral proteases (Chapter 1, fig. 12 and 13, pages 61-62). Furthermore it could be constructed without requiring new SIV PCR primers, or necessitating PCR amplification of the UDG gene to engineer a suitable site at its 3' end, or the need to use cassettes or complex rounds of mutagenesis.

The strategy with respect to construction of the UDG-PR junction was intended to maximise the possibility of autocatalysis whilst simplifying plasmid construction. It achieved both goals.

Isolation of the desired UDG-SIV fusion construct from miniprep DNA.

Transformants were screened (as earlier described) by digestion with NcoI (fig. 12), amplification of the SIV_{AGM} PR gene by PCR (fig. 13), and by direct sequencing (fig. 14). Only clone 6. was positive in all respects, and this was designated $pMIRA_6$.

Figure 12 NcoI screens.



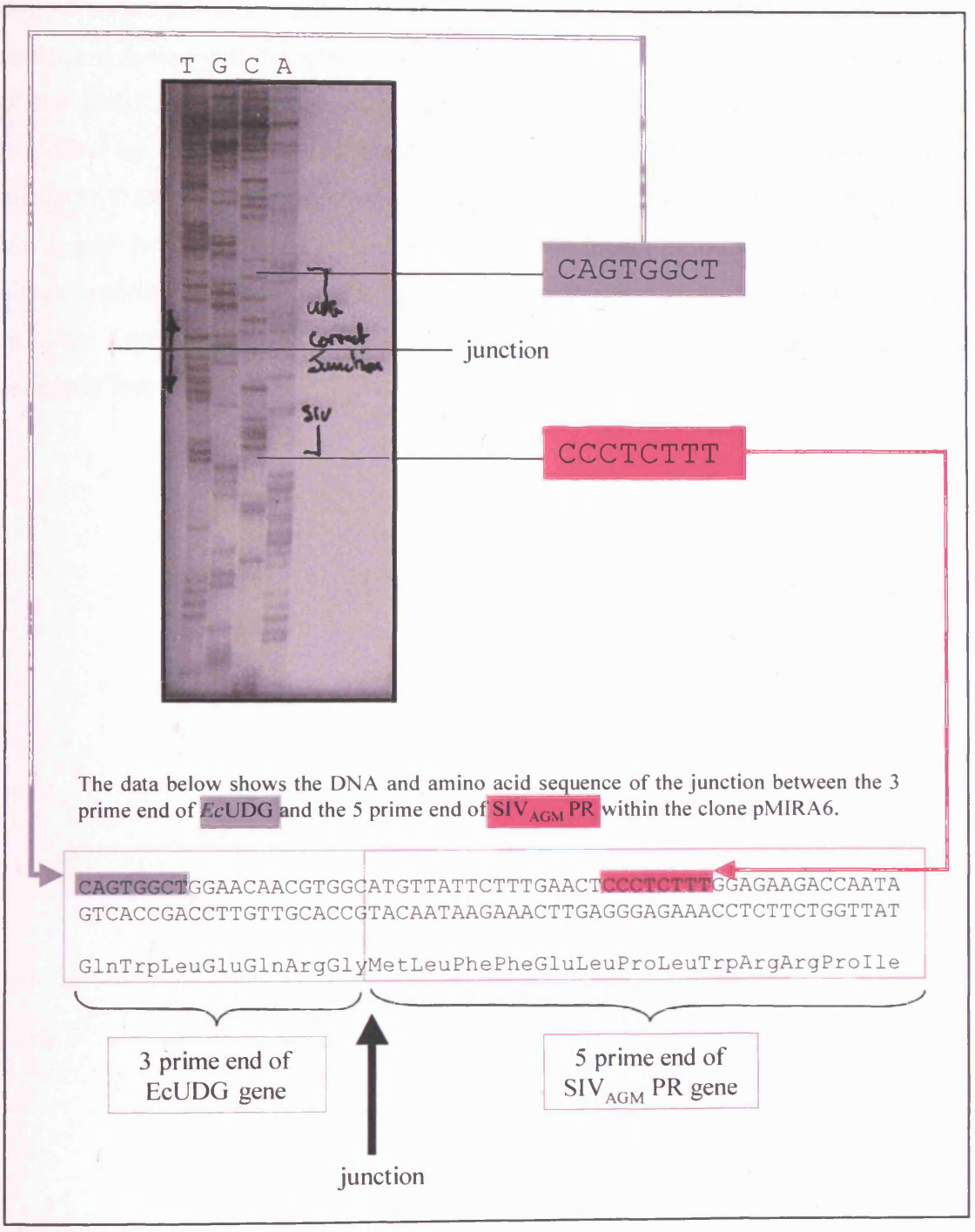
Positive clones released a 0.78 kb fragment (indicated by arrow) on digestion with NcoI.

Figure 13 PCR screens.

← SIV_{AGM} PR PCR product
← Primer dimer product

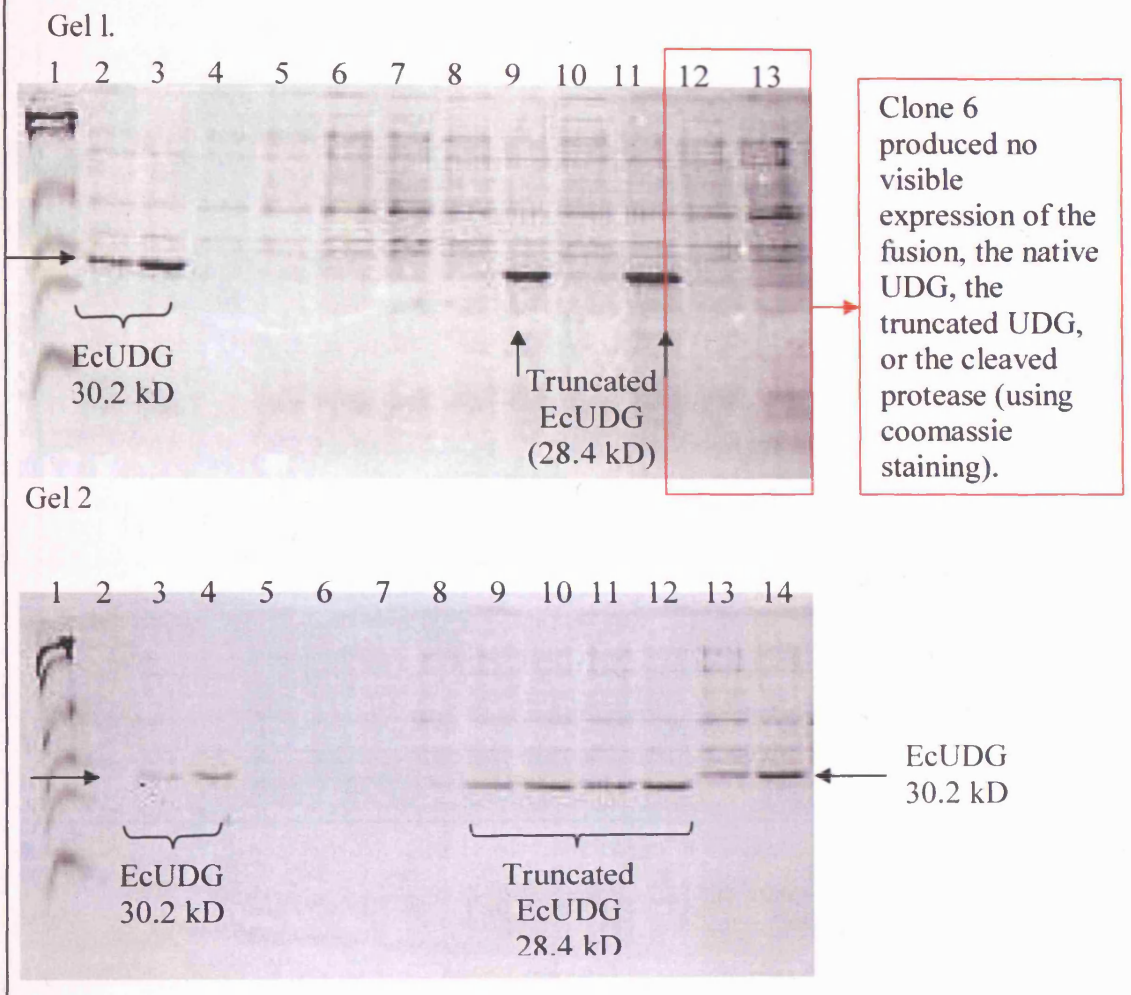
Positive clones were used as templates in a PCR reaction with SIV_{AGM} PR primers, and those producing an amplified product were selected for sequencing.

Figure 14 Sequence across *Ec*UDG-SIV$_{AGM}$ PR junction in pMIRA6.



Small-scale bacterial expression of the UDG-SIV fusion protein from pMIRA$_6$.
The small-scale expression screen showed that pMIRA$_6$ did not overexpress the desired fusion product at high levels (fig. 15, gel 1, lanes 12 and 13). The full-length, uncleaved fusion was detected by western blot (fig. 16) within the insoluble fraction of the pMIRA6 cell lysate using antibodies raised against the 6 x Histidine tag (supplied by Clonetech), and small scale IMAC experiments (not shown) on the insoluble fraction confirmed that the fusion was expressed and could be isolated at low levels. No evidence of a truncated or cleaved fusion product was detected during either experiment. These results suggested that the protease was not cleaving itself from the fusion within the bacterial cells, yet expression levels of the fusion product remained low despite this.

Figure 15 Small scale expression screens of the UDG-SIV (pMIRA) clones.

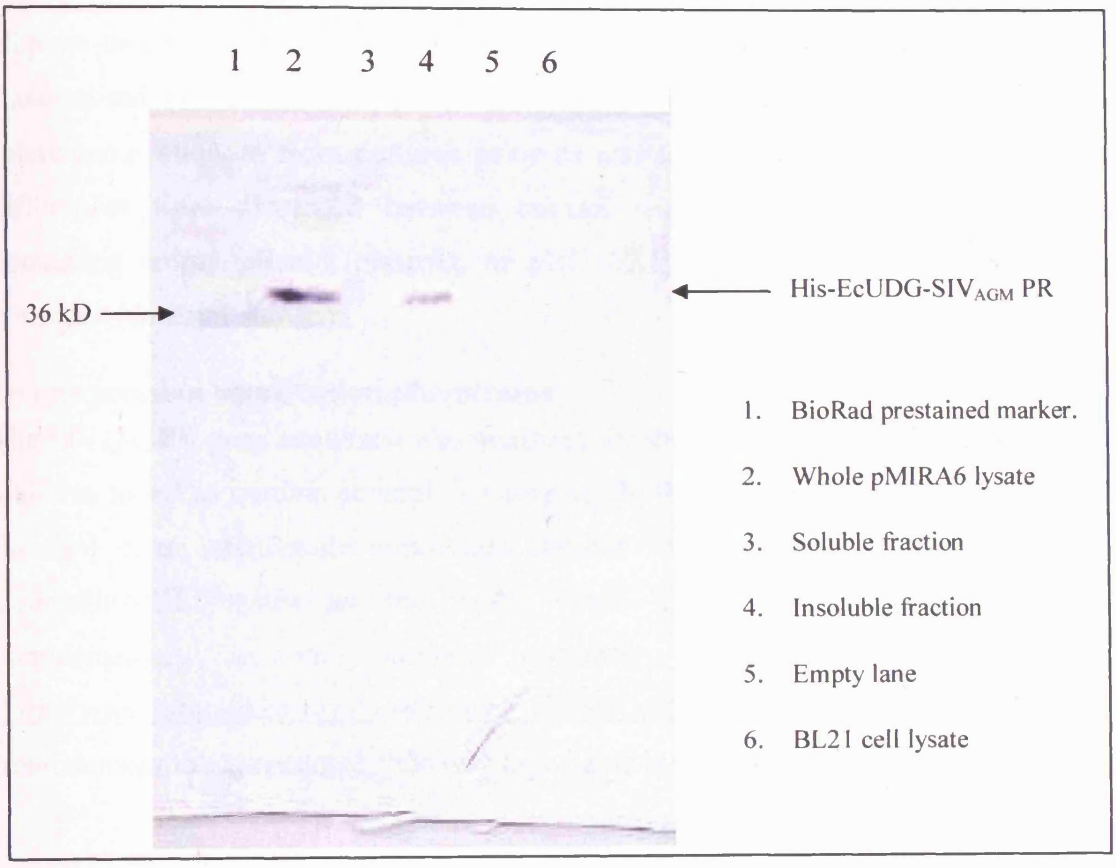


Gel 1		Gel 2	
1	Low range prestained BioRad standards	1	Low range prestained BioRad standards
2 / 3	pBU107.1 uninduced / induced	2	empty lane
4 / 5	pRsetB uninduced / induced	3 / 4	pBU107.1 uninduced / induced
6 / 7	BL21 (DE3) uninduced / induced	5 / 6	pRsetB uninduced / induced
8 / 9	Clone 1 uninduced / induced	7 / 8	BL21 (DE3) uninduced / induced
10 / 11	Clone 4 uninduced / induced	9 / 10	V + L 1 uninduced / induced
12 / 13	Clone 6 uninduced / induced	11 / 12	V + L 5 uninduced / induced
		13 / 14	-ve clone 2 uninduced / induced

V + L (vector+ligase) = cut and religated vector (pBU107.1).

pBU107.1 encoded *Ec*UDG. pRsetB = parent vector.

Figure 16 Western blot of small scale pMIRA6 expression screen.



The possibility that the fusion itself was toxic was investigated by comparison of cell growth rates (as assessed by measuring optical densities at λ = 600 nm for 1 ml aliquots of cultures at regular time intervals), by transformation efficiencies (using standardised DNA and competent cell preparations), and using viable cell counts (plating out aliquots from cultures prior to and following IPTG induction) but no differences were discerned between control strains (either without plasmid, or containing empty pRsetB plasmid, or pBU107.1) and cells containing the fusion construct (data not shown).

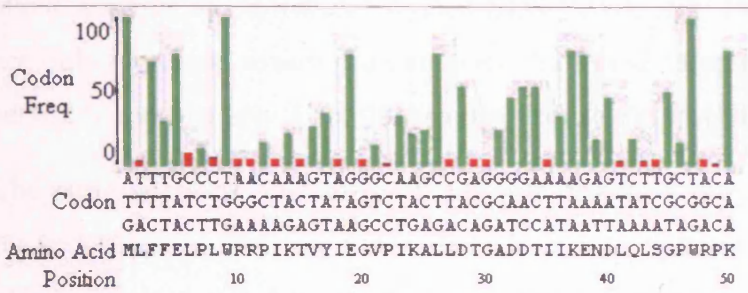
Overexpression using codon plus strains

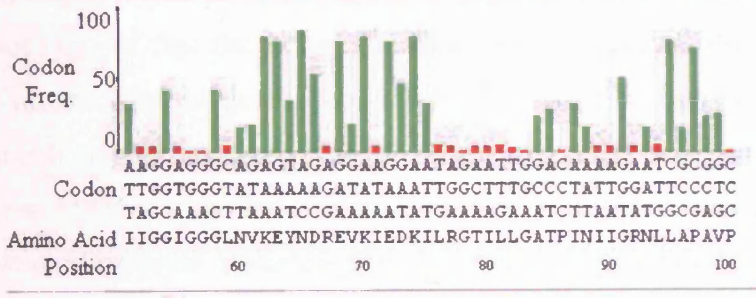
The SIV_{AGM} PR gene sequence was analysed for the presence of rare codons (fig. 17) and was found to contain several. To investigate the influence of rare codons within the SIV gene, small-scale expression screens were carried out using the BL21-CodonPlus-RIL strain as the host. These cells express additional tRNAs complementary to rare codons encoding Arg, Isoleucine and Lysine (http://www.stratagene.com/pdf/brochure/BR38.pdfande=7317). Suprisingly, no improvement in expression levels was detectable by PAGE (data not shown).

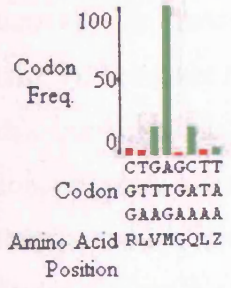
Figure 17 Codon usage analysis of SIV_{AGM} PR gene sequence.

***E. coli* Codon Usage Analysis 2.0** by Morris Maduro **Sequence**: SIV_{AGM} PR (324 bases)

```
5'-ATGTTATTCT TTGAACTCCC TCTTTGGAGA AGACCAATAA AGACAGTGTA TATAGAAGGG GTCCCCATTA AGGCACTGCT
   AGACACAGGG GCAGATGACA CCATAATTAA AGAAAATGAT TTACAATTAT CAGGTCCATG GAGACCCAAA ATTATAGGGG
   GCATAGGAGG AGGCCTTAAT GTAAAAGAAT ATAACGACAG GGAAGTAAAA ATAGAAGATA AAATTTTGAG AGGAACAATA
   TTGTTAGGAG CAACTCCCAT TAATATAATA GGTAGAAATT TGCTGGCCCC GGCAGTGCCC CGGTTAGTAA TGGGACAATT
   ATAA-3'
```







Colors: ■= less than 10% of codons for same amino acid; ■= at least 10%

Fraction of sense codons below threshold (=10.00): 45/107 (42%)

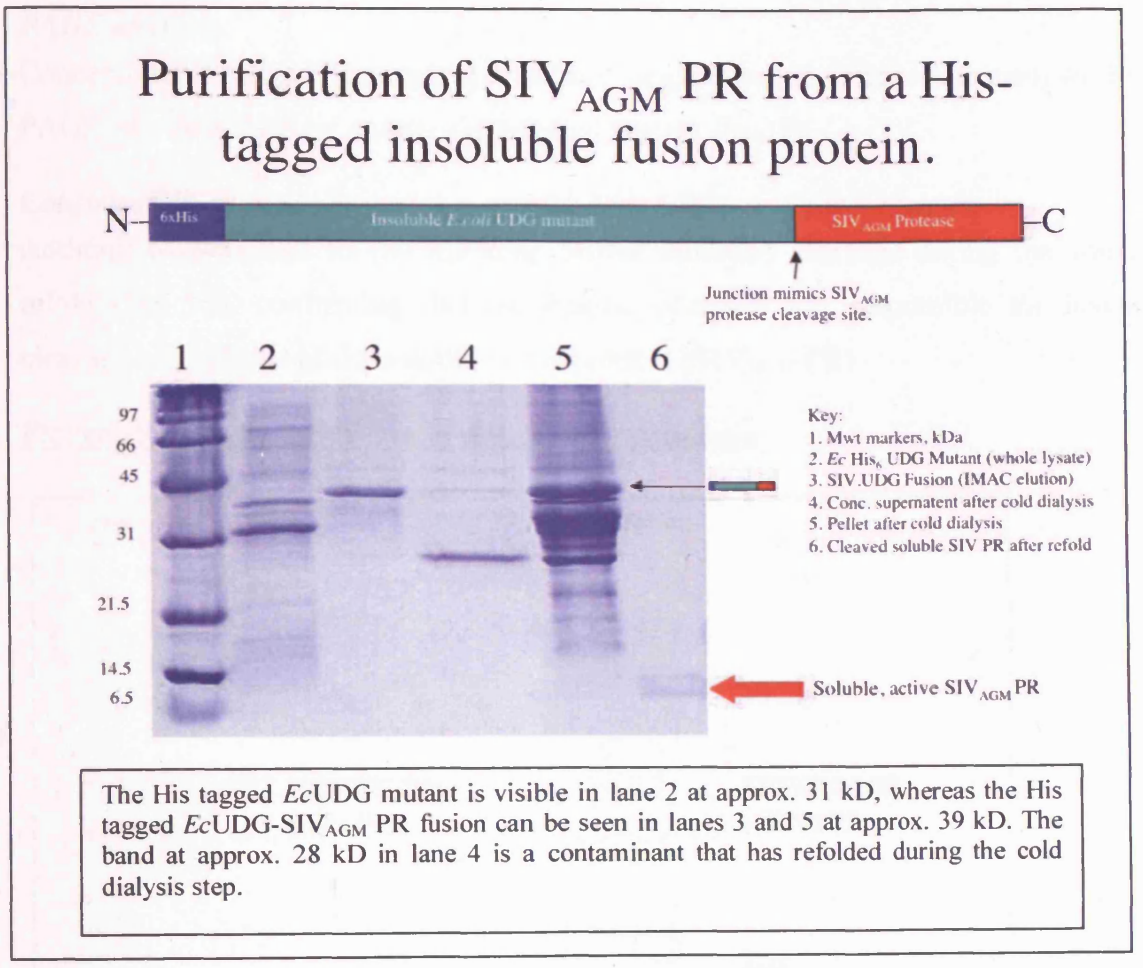
The above figure (17) plots the frequency with which codons of the SIV_{AGM} PR gene are found in *E.coli*. Those codons used by *E.coli* on less than 10% of possible occasions are plotted in red.

Large scale expression and purification of the EcUDG-SIVAGM PR fusion.

Since it had been established that a fusion product was expressed, albeit at low levels, it was decided to pursue a strategy of large-scale cell culture in order to obtain working quantities of the fusion protein. Protein was prepared from multiple large-scale (8 litre) cultures, as described in the methods section of this chapter. Aliquots from different stages of the preparation were concentrated by acetone precipitation to allow analysis by coomassie stained SDS-PAGE (fig. 18). The acetone precipitation not only increased fusion protein concentration to detectable levels for the gel, it also served to remove urea from the sample prior to electrophoresis.

The purpose of the cold dialysis step was primarily to concentrate the fusion protein by removing the chaotrope. The dialysis was carried out at low temperature and at a pH that would not be expected to support aspartic (and hence SIV_{AGM}) protease activity, so that the protease would remain fused to the insoluble UDG throughout. The precipitated fusion protein was harvested by centrifugation and redissolved in a much smaller volume of chaotropic buffer, thus concentration was achieved (fig. 18 lane 5). This step also ensured that other insoluble contaminant proteins able to refold were eliminated in the discarded supernatent following the cold dialysis step (fig. 18 lane 4). The fusion product appeared to be essentially pure following IMAC (fig. 18 lane 3), however the cold dialysis step greatly concentrated the sample and following this numerous contaminants could be seen (fig. 18 lane 5). However, since these contaminants were insoluble, and had failed to refold during the cold dialysis step, they posed no problems in terms of contamination of the final SIV_{AGM} PR solution. The concentration of the sample by cold dialysis also ensured a greater density and proximity of refolded protease molecules during the warm dialysis, increasing the likelihood of dimerisation, activity and autocatalysis. The SIV_{AGM} PR released during the warm dialysis step remained in solution, and was consequently diluted, but it was still clearly visible on a coomassie stained gel after concentration by acetone precipitation. No contaminant bands were visible in the final preparation (fig. 18 lane 6).

Figure 18 Large scale preparation of the EcUDG-SIV$_{AGM}$ PR fusion product.



The His tagged *Ec*UDG mutant is visible in lane 2 at approx. 31 kD, whereas the His tagged *Ec*UDG-SIV$_{AGM}$ PR fusion can be seen in lanes 3 and 5 at approx. 39 kD. The band at approx. 28 kD in lane 4 is a contaminant that has refolded during the cold dialysis step.

Characterisation of the active, refolded, SIV_{AGM} protease

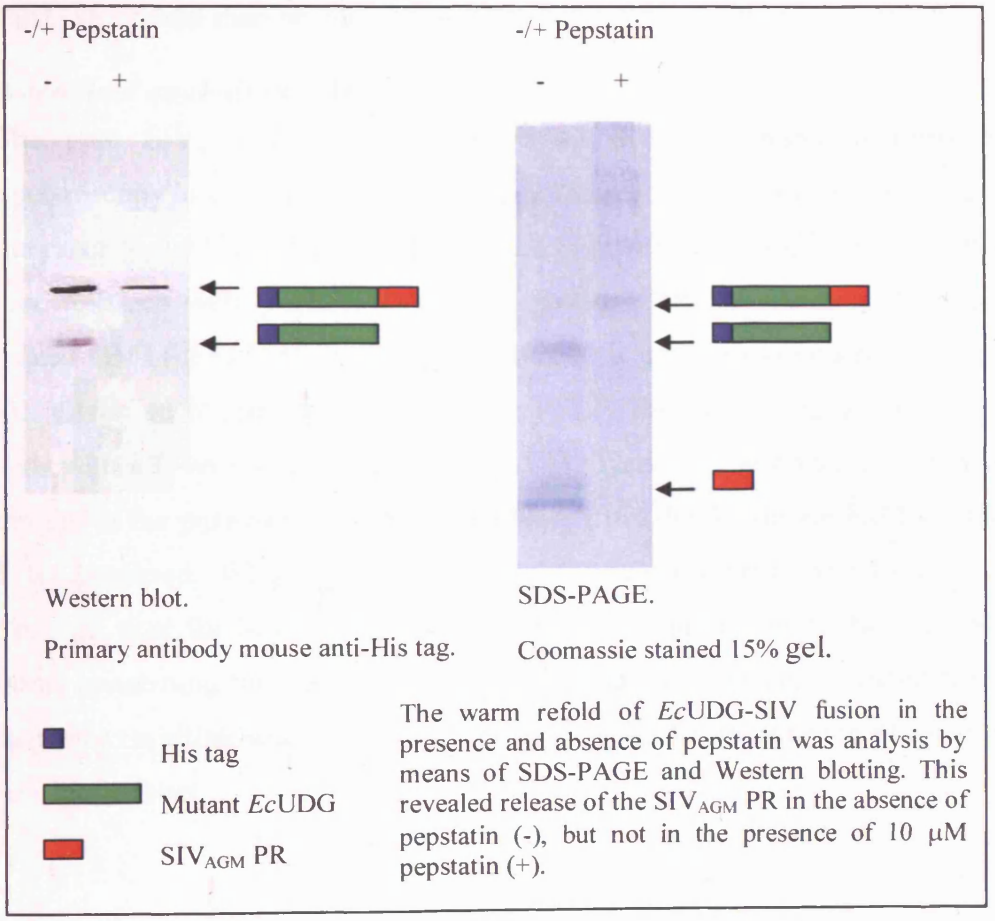
PAGE analysis.

Concentrated SIV_{AGM} PR appeared as a pure single band of 11 kD when analysed by PAGE and visualised by coomassie staining (results fig. 18).

Confirmation of aspartic protease activity by inhibition of autocatalysis.

Addition of pepstatin to the refolding buffer inhibited cleavage during the warm refold (fig. 19), confirming that an aspartic protease was responsible for fusion cleavage and release of the soluble 11 kD product (SIV_{AGM} PR).

Figure 19 The inhibition of cleavage by pepstatin.



Immunological characterisation of SIV_{AGM} PR.

Dot blots were carried out to investigate the possibility that SIV_{AGM} PR might be immunologically similar to HIV-1 or HIV-2 PR. Primary antibodies raised against the HIV proteases cross-reacted only very weakly with the SIV_{AGM} PR. These antibodies could not therefore reasonably be used for future detection of the SIV_{AGM} PR.

Crystallographic analysis of SIV_{AGM} PR

No crystallisation was observed under the conditions tested, however, these were not comprehensive, and further screens would have been desirable. Future preparations of the enzyme could be used to carry out sparse matrix screens in order to ascertain the preferred conditions for SIV_{AGM} PR crystallisation, and optimisation of these conditions could then be carried out.

N-terminal analysis and Mass Spectroscopy.

The pure SIV_{AGM} PR was subjected to both N-terminal analysis and Mass Spectroscopy in order to provide both a positive identity of the protein, and to identify the exact N and C terminal cleavage sites. N-terminal analysis confirmed that the first nine residues were FELPLWRRP, suggesting that the fusion's cleavage site was indeed GMLF:FELP as predicted. MS analysis gave a monomeric mass of 11142.0 kD. Given an N-terminus beginning at FELP, the mass suggests that the monomer ends with a C-terminal sequence of RLVM. Thus we can conclude that the N and C termini of the pure SIV_{AGM} PR after release from the fusion are N-FELP and RLVM-C as predicted. Whilst this does not absolutely confirm the nature of the native cleavage sites for SIV_{AGM} PR, it does strongly support the predictions made in this thesis concerning their location. A similar construct utilising extended regions of the sequence on either side of the protease gene could be produced in order to provide an absolute answer.

Why was expression of the EcUDG-SIV$_{AGM}$PR fusion so poor?

Taken in context with subsequent sequencing data, the poor expression levels of the UDG-SIV PR fusion protein from clone 6 suggested several possibilities.

First, perhaps the SIV-PR was able to cleave itself from the UDG fusion (possibly because the truncation of the UDG followed by addition of the SIV had rendered the UDG soluble once more), and was thus able to exert a toxic effect on the cells. However, the truncated UDG was already known to be insoluble. It was predicted that the fusion would be generally hydrophobic in nature so that it should also be insoluble. Furthermore, the fusion protein expressed was always detected exclusively in the insoluble fraction following cell lysis, and no protease activity was ever detected in the soluble fraction.

If the protease had been able to cleave itself from the fusion *in vivo* we would expect to see the formation of at least two bands on induction, one corresponding to the protease, the other to the truncated UDG. However, no such bands were observed. This could be explained if the protease, having released itself from the fusion, was able to cleave the UDG into numerous fragments too small to resolve. If release of the protease occurred shortly after induction, it may also have exerted a cytotoxic effect further limiting production of the fusion. As a result no overexpressed bands (fusion, truncated UDG, or released PR) would be visible on the gel, whilst relatively small amounts of the protease activity would be present. However, the incubation of mutant UDG with the HIV proteases suggested that the insoluble form of UDG was resistant to proteolysis by these enzymes, and therefore probably also resistant to cleavage by SIV proteases. Furthermore, the limited expression that is observed occurs in the form of an uncleaved fusion suggesting that the *Ec*UDG is not cleaved in the host cell. Finally the UDG portion of the fusion product remains in tact even after autocatalytic release of the SIV$_{AGM}$ PR in optimised buffer at 37°C (as is illustrated by fig. 19), this strongly supports the idea that the mutant EcUDG is resistant to cleavage by SIV$_{AGM}$ PR. Therefore degradation of the *Ec*UDG by the PR is unlikely, and thus cannot explain the absence of bands corresponding to overexpressed *Ec*UDG or fusion from analyses of lysates by SDS-PAGE.

Potentially the fusion itself may be toxic - despite being insoluble, inactive, and composed primarily of a non-toxic *E.coli* protein that is ordinarily highly overexpressed. There is a possibility that the fusion was somehow interupting the function of the host cell's own UDG or related proteins. If this were the case, however, we would not expect to see the mutant UDG overexpressed – whereas in fact it is normally highly expressed apparently without deleterious effect. Also given that the protease is toxic by virtue of its proteolytic activity, and given that if active it would release itself from the fusion, we can assume that the fusion itself is non-toxic with respect to protease activity. This is the premise for the fusion design. The possibility that the fusion itself was toxic was nevertheless investigated by comparison of cell growth rates, transformation efficiencies and viable cell counts, yet no difference in the relative vitalities of bacteria with or without $pMIRA_6$ (the UDG-SIV PR fusion vector) prior to or following induction, was observed (data not shown).

It is interesting to compare the expression of UDG-SIV PR fusion with the HIV-2 PR active site mutants described in Chapter 6. The expression levels of HIV-2 PR increased hugely once the active site was mutated. This cannot be explained in terms of codon usage, since it represents only a minimal change to the DNA sequence. It is instead presumably the result of knocking out proteolytic activity and thus reducing the deleterious effect of the protease on the host cells. It is strange then that, given the fusion similarly renders SIV inactive, we do not see a significant increase in SIV yield. A comparison of codon usage in the HIV-2 PR and SIV_{AGM} PR genes reveals that the HIV-2 PR gene sequence contains fewer rare codons than the SIV_{AGM} PR sequence, and HIV-2 PR does not suffer the long runs of consecutive rare codons found in the SIV gene (see fig. 20 for analysis of HIV-2 PR codon usage).

Abundance of tRNA and codon usage

Finally, rare codons could be the source of low protease yields, and consequently of low fusion yields.

The genetic code contains 61 codons encoding 20 amino acids (Crick *et al.* 1961). This means that a given amino acid may be encoded by several distinct (yet synonymous) codons, each read by one or more transfer RNAs (tRNAs). Where several codons encode the same amino acid they do not necessarily arise with the same frequency within a given organism's genome (Nakamura 1996), or within its' population of messenger RNAs (mRNAs) (Dong 1996). In fact certain codons are

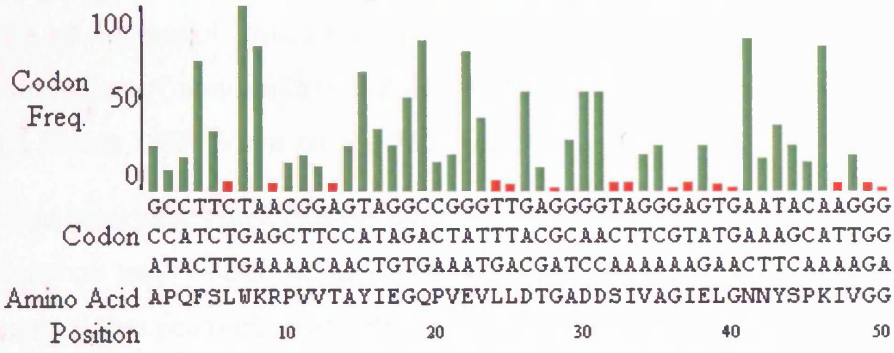
apparently favoured over others encoding the same amino acid - a phenomenon known as codon bias - and the precise nature of this bias varies between species. The frequency of a codon's use generally correlates with the availability of the tRNA that recognises it (or *vice versa*) (Dong 1996, Ikemura 1981), and in highly expressed genes the most frequently used codons tend to be those recognised by the most abundant tRNA (Post *et al.* 1979). It has been demonstrated that this arrangement favours more rapid protein synthesis and lower rates of amino acid substitution errors than would otherwise be observed (Ehrenberg and Kurland 1984).

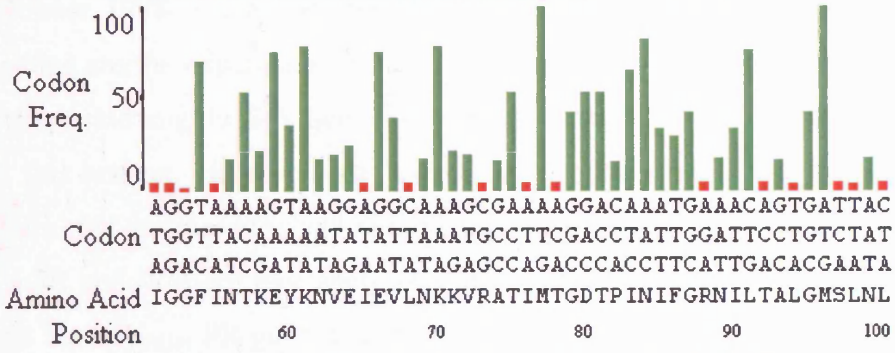
Figure 20 Codon usage analysis of the HIV-2 PR gene sequences.

***E. coli* Codon Usage Analysis 2.0** by Morris Maduro

Sequence: HIV-2 PR (300 bases)

```
5'-GCACCTCAAT TCTCTCTTTG GAAAAGACCA GTAGTCACAG CATACATTGA GGGTCAGCCA
   GTAGAAGTTT TGTTAGACAC GGGAGCTGAC GACTCAATAG TAGCAGGAAT AGAGTTAGGA
   AACAATTATA GCCCAAAAAT AGTAGGGGGA ATAGGGGGAT TCATAAATAC CAAGGAATAT
   AAAAATGTAG AGATAGAAGT TCTAAATAAA AAGGTACGGG CCACCATAAT GACAGGCGAC
   ACCCCAATCA ACATTTTTGG CAGAAATATT CTGACAGCCT TAGGCATGTC ATTAAATCTA -3'
```





Colors: ■= less than 10% of codons for same amino acid; ■= at least 10%

Fraction of sense codons below threshold (=10.00): 30/100 (30%)

Given the correlation between frequency of codon usage and tRNA abundance, plus the observation that rapid and high level protein expression occur where the encoding mRNA contains frequently used codons, it is reasonable to expect that the opposite would also be true. Where infrequently used codons (rare codons) predominate in an mRNA its' translation would be predicted to be problematic. A low rate and level of protein expression and an increased rate of substitution errors might be expected, and has been observed (Robinson *et al.* 1984, Kane 1995). A comparison of protein synthesis involving different codons in mutants of the *E.coli lacz* gene showed that rare codons were translated at a rate of only 2 peptide bonds/sec, equivalent to about 20% of the rate of common codons (Sorensen, *et al.* 1989). Other work supports this observation of non-uniform translation rates at different codons in *E.coli* (Bonekamp and Jensen 1988, Sorensen and Pedersen 1991).

It had been anticipated that the position of the SIV gene following such a large fusion sequence may overcome any potential codon usage problems, since it has been reported that positioning of rare codons can influence their effect on expression levels (Bulmer 1988; Chen and Inouye 1994; Goldman, Rosenberg *et al.* 1995). If rare codons are the explanation for the persistently low levels of SIV_{AGM} PR expression, then positioning the SIV gene downstream of the UDG did not have the desired effect in this respect. However, the use of codon plus strains, and in particular the RIL strain, did not produce any significant improvement in yield (as assessed by SDS-PAGE and Western blot analyses – data not shown) either. This seemed to indicate that the SIV_{AGM} PR gene codon usage is so incompatible with *E.coli* that codon plus strains are insufficient to overcome the problem.

Investigation of the SIV_{AGM} PR sequence using the Codon Usage Analysis 2.0 program (By Morris Maduro) revealed that 42% of the SIV codons occur at a frequency of less that 10% in *E.coli* (fig. 17), while 36% occur at a frequency of less than 5%. Using the term "rare codon" as arbitrarily defined by Wada *et al.* (Wada, Wada *et al.* 1992) (and susbequently used by Kane 1995 (Kane 1995)), to mean codons used at a frequency of less than 1%, then 15% of the SIV_{AGM} PR gene codons are rare in *E.coli*. In particular there are seven rare Arginine codons (six are AGA - which occur at a frequency of 2.1 per 1000 codons in *E.coli* - and 1 is CGG which occurs at a frequency of 4.6 per 1000 *E.coli* codons) including an adjacent pair located near the start of the gene.

The analysis also confirmed the presence of rare codons for Threonine, Glycine, Proline and Serine in addition to the rare Arginines, Isoleucines and Leucines that were compensated for by the three rare tRNAs expressed by the RIL strains. The presence of these non-RIL rare codons explains the ineffectiveness of the RIL strains. Of the 45 most rare codons 16 were non-RIL. Codons occurring at a frequency of less than 10% were distributed throughout the gene such that there were seven instances where two were adjacent, one instance of three being adjacent, and one instance of eight being adjacent. The row of eight suboptimal codons occurred in the middle of the gene and included 2 that are classified as rare according to the Wada definition.

The occurence in a gene of two or more adjacent rare codons is known to have a deleterious effect on protein expression (Rosenberg, Goldman *et al.* 1993; Kane 1995; Phoenix, Pewsey *et al.* 1995), consequently the presence of a run of eight rare codons would presumably prove extremely difficult for a host cell to express.

Given that the large number of rare codons is the likely cause of poor SIV_{AGM} PR expression in *E.coli* there are perhaps two possible solutions, change the host organism or synthesis an optimised gene. Possible alternative hosts may include *Bacillus subtilis*, which has no codon preference (Ogasawara 1985; Loshon, Tovar-Rojo *et al.* 1989) and has also been used for protein expression (Conrad, Savchenko *et al.* 1996; Chan, Chan *et al.* 2002). Possibly eukaryotic host cell based systems (eg. yeast, insect, or mammalian) would provide viable alternatives, but are typically more expensive and technically more difficult than their bacterial counterparts. Total synthesis of the protease gene using codons optimised for expression by *E.coli* is preferable since it would allow the use of established expression methods and *E.coli* as the host. The optimisation of codon usage by gene synthesis has successfully augmented the expression of other proteins (Graham, Atkinson *et al.* 1993; Martin, Vrhovski *et al.* 1995) including the korB protein from Pseudomonas in our laboratory (S.Singh personal communication). One possible caveat is that there is probably a limit to the size of gene that can be successfully synthesised by current methods, and in such a situation an alternative host organism may prove the better alternative unless mutagenesis of key rare codons was sufficient to restore expression levels.

Given the results of this study it was considered that poor expression of the fusion very probably resulted from the widespread presence of rare codons within the

SIV_{AGM} PR gene sequence. To establish the validity of this suggestion would require the application of one of the solutions discussed above. The SIV_{AGM} PR gene is sufficiently short that synthesis of the optimised gene is quite feasible, whilst conventional mutagenesis would require multiple reactions to be carried out in order to eliminate the runs of adjacent rare codons. Consequently gene synthesis would be the preferred approach.

The 'Recursive PCR' method of Prodromou and Pearl (Prodromou and Pearl 1992) would have been the method of choice, though financial and practical constraints (ie. the cost of the large oligos required, and the need for a computer program specifically to design them) precluded this experiment.

For the purposes of discussion the procedure for design of recursive PCR primers would have been as follows:-

1. Carry out codon usage analysis of SIV_{AGM} PR with respect to frequency of codon occurrence in *E.coli* genes (fig. 17), and locate any rare codons.

2. Generate a modified SIV_{AGM} PR gene sequence (eg. using GCG, Genetics Computer Group, now Accelrys) such that all codons are now optimised and found with high frequency in *E.coli*. Ensure that the amino acid sequence has not been inadvertently altered in the process.

3. Map the restriction sites present in the optimal gene, and choose useful, unique sites (preferably at regular intervals - eg 30 bp intervals) - or if necessary alter codons to create such sites - which may be used to clone parts of the synthetic gene if it proves impossible to make the whole gene in one piece.

4. Select restriction sites to be incorporated at the 5' and 3' ends of the gene for cloning purposes, and design suitable sequences for use in the exterior primers (ensuring that the gene will be inserted in frame once cloned). This requires knowledge of the vectors you may wish to use in future.

5. From the sequence thus far created, design a set of overlapping primers, which alternate between the positive (5' to 3') strand, and the negative (3' to 5') strand. These should preferably be about 60 bps in length and should overlap with their adjacent primers by 10-15 bps.

6. The primer design must now be adjusted to ensure that the Tm of each overlap is approximately the same as the Tm for all the others, preferably they should all have a value of about 60°C. Varying the respective lengths of the primers may be required at this point to obtain the desired Tm values. The position of unique restriction sites should also be checked to ensure that they are usefully located with respect to the primers.

7. This first generation of primer designs now have to be checked relative to each other to ensure that they would not mis-prime, and will only anneal to the intended complementary primer. Where mis-priming is detected the primers involved must be redesigned by modifying the sequence as required to prevent mis-priming, but while maintaining optimal codons as far as possible. In this way a second generation of primer designs is produced, and these are once more checked against each other to ensure there is no mis-priming. This cyclic checking and modifying process is repeated until an optimal design is achieved. The primers can then be synthesised.

This last cyclical step is critical since mis-priming is highly likely if the design is poor, and it will result in a total failure of the reaction. In reality this step is usually too complex to be carried out manually and all successful use of recursive PCR has relied on computer aided primer design, with programs written specifically for the task of checking recursive PCR primers against each other. Until recently this meant that experimentalists had to write their own program, however, software (Juniper) has now been developed for the purpose (C. Richardson, unpublished). Juniper accepts the amino acid sequence in FASTA format and produces an optimised nucleotide sequence, from which it designs appropriate primers. This was used to design primers (figs. 21 and 22) for synthesis of an optimised SIV_{AGM} PR gene (fig. 23) in the future.

Figure 21 Results from Juniper.

The SIVagm PR gene was optimised for expression in *E.coli*, and designed to incorporate a 5' AflIII site, a 3' HindIII, and a Stop codon. The SIV amino acid sequence encoded begins with the complete N-terminal PR recognition sequence. The resultant nucleotide sequence was then used to generate overlapping oligos, as shown below.

Top strand: 5' to 3'

```
GCCGCGCACATGCGTGGCATCTTCTTCGAACTGCCGCTGTGGCGTCGTCCGATCAAG
ACCGTTTACATCGAAGGTGTTCCGATTAAAGCTCTGCTGGACACCGGTGCTGACGAC
ACCATCATCAAAGAAAACGACCTGCAGCTGTCCGGTCCGTGGCGTCCGAAAATCATC
GGCGGTATCGGTGGTGGTCTGAACGTTAAAGAATACAACGACCGTGAAGTTAAAATC
GAAGACAAAATCCTGCGTGGTACCATCCTGCTGGGTGCTACCCCGATCAACATCATC
GGTCGTAACCTGCTGGCTCCGGCTGTTCCGCGTCTGGTTATGGGTCAGCTGTAAGCT
TGGCGC
```

Bottom strand: 5' to 3'

```
GCGCCAAGCTTACAGCTGACCCATAACCAGACGCGGAACAGCCGGAGCCAGCAGGTT
ACGACCGATGATGTTGATCGGGGTAGCACCCAGCAGGATGGTACCACGCAGGATTTT
GTCTTCGATTTTAACTTCACGGTCGTTGTATTCTTTAACGTTCAGACCACCACCGAT
ACCGCCGATGATTTTCGGACGCCACGGACCGGACAGCTGCAGGTCGTTTTCTTTGAT
GATGGTGTCGTCAGCACCGGTGTCCAGCAGAGCTTTAATCGGAACACCTTCGATGTA
AACGGTCTTGATCGGACGACGCCACAGCGGCAGTTCGAAGAAGATGCCACGCATGTG
CGCGGC
```

Oligo overlaps:

```
  1 GCCGCGCACATGCGTGGCATCTTCTTCGAACTGCCGCTGTGGCGTCGTCC
    .................................................

 51 GATCAAGACCGTTTACATCGAAGGT.........................
    ......CTGGCAAATGTAGCTTCCACAAGGCTAATTTCGAGACGACCTGT

101 ..........CGACACCATCATCAAAGAAAACGACCTGCAGCTGTCCGGT
    GGCCACGACTGCTGTGGTAGTAGTTTCTTT...................

151 CCGTGGCGTCCGAAAATCATCGGCGGTATCGGTGG..............
    ..................TAGCCGCCATAGCCACCACCAGACTTGCAATT

201 ....................TTAAAATCGAAGACAAAATCCTGCGTGGTA
    TCTTATGTTGCTGGCACTTCAATTTTAGCTTCTGTTTTAGGA........

251 CCATCCTGCTGGGTGCTACCCCGATCAACATCATCGGTCGTAACC.....
    ..........................TTGTAGTAGCCAGCATTGGACGAC

301 .................................................
    CGAGGCCGACAAGGCGCAGACCAATACCCAGTCGACATTCGAACCGCG
```

Figure 22 Results from Juniper.

The oligonucleotide designs thus generated were as shown below.

Forward oligos: 5' to 3'

1.GCCGCGCACATGCGTGGCATCTTCTTCGAACTGCCGCTGTGGCGTCGTCCGATCA
AGACCGTTTACATCGAAGGT

2.CGACACCATCATCAAAGAAAACGACCTGCAGCTGTCCGGTCCGTGGCGTCCGAAA
ATCATCGGCGGTATCGGTGG

3.TTAAAATCGAAGACAAAATCCTGCGTGGTACCATCCTGCTGGGTGCTACCCCGAT
CAACATCATCGGTCGTAACC

Reverse oligos: 5' to 3'

1.TTTCTTTGATGATGGTGTCGTCAGCACCGGTGTCCAGCAGAGCTTTAATCGGAAC
ACCTTCGATGTAAACGGTC

2.AGGATTTTGTCTTCGATTTTAACTTCACGGTCGTTGTATTCTTTAACGTTCAGAC
CACCACCGATACCGCCGAT

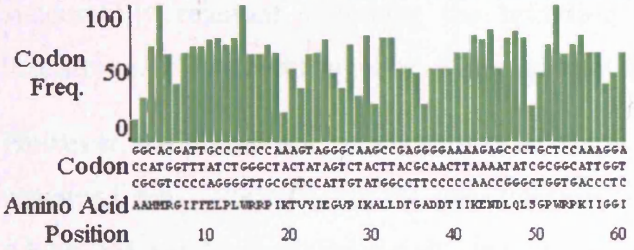
3.GCGCCAAGCTTACAGCTGACCCATAACCAGACGCGGAACAGCCGGAGCCAGCAGG
TTACGACCGATGATGTT

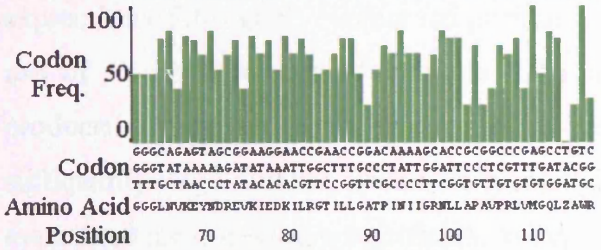
Figure 23 Codon analysis of the proposed synthetic SIV_{AGM} PR gene.

***E. coli* Codon Usage Analysis 2.0** by Morris Maduro

Sequence entered (348 bases):

```
5'-GCCGCGCACA TGCGTGGCAT CTTCTTCGAA CTGCCGCTGT GGCGTCGTCC GATCAAGACC GTTTACATCG AAGGTGTTCC
   GATTAAAGCT CTGCTGGACA CCGGTGCTGA CGACACCATC ATCAAAGAAA ACGACCTGCA GCTGTCCGGT CCGTGGCGTC
   CGAAAATCAT CGGCGGTATC GGTGGTGGTC TGAACGTTAA AGAATACAAC GACCGTGAAG TTAAAATCGA AGACAAAATC
   CTGCGTGGTA CCATCCTGCT GGGTGCTACC CCGATCAACA TCATCGGTCG TAACCTGCTG GCTCCGGCTG TTCCGCGTCT
   GGTTATGGGT CAGCTGTAAG CTTGGCGC-3'
```





Colors: ■ = less than 10% of codons for same amino acid; ■ = at least 10%

Fraction of sense codons below threshold (=10.00): **0/115 (0%)**

SUMMARY

A novel expression system was developed (based on a pRsetB based construct containing a mutant of the *E.coli* UDG) allowing the expression of an insoluble UDG-SIV_{AGM} PR fusion protein as inclusion bodies. The system was designed to eliminate the cytotoxic effects of the protease by rendering it insoluble. It was also intended to provide a simple two step purification of the protease by means of IMAC, and autocatalytic release of the protease under specific refolding conditions. The presence of the construct had no apparent toxic effect on host cells, even during induction. The fusion was expressed insolubly, purified in a single IMAC step, and the protease was successfully released following the refolding procedure and separated from the insoluble UDG by centrifugation. The construct design was thus considered a success.

However, the fusion was poorly expressed despite lack of toxicity, and protease yields remained low. The SIV_{AGM} PR gene sequence contains a large number of codons which are rarely used by *E.coli*, and this probably explains the persistently low expression of this gene. Neither the position of the gene downstream of UDG, nor the use of Codon Plus strains (such as RIL) in conjunction with the fusion system produced any improvement in expression. Presumably the SIV sequence contains a sufficiently large number, wide variety, and clustered distribution of rare codons that even these measures were insufficient to overcome the problem. It was concluded that the best solution would be total synthesis of the protease gene using codons optimised for expression by *E.coli*, and a method was proposed.

ACKNOWLEDGEMENTS

Thanks are due Prof. Masanori Hayami for provision of the original SIV_{AGM} PR DNA template, to Andrew Shearer for construction of the pACS28 pMALcR1 based plasmid, and to Renos Savva for the gift of plasmid pBU107.1 plus our many useful discussions. Thanks to Chris Prodromou, Chris Van der Waals, Mark McAllister, Shradha Singh, and Chris Richardson for their advice with respect to codon usage problems and recursive PCR methodology.

Chapter Seven: General Discussion.

This thesis has described the cloning, expression, and purification, of three native retroviral proteases from HIV-1, HIV-2 and SIV_{AGM}. Also reported were the crystallisation and preliminary crystallographic analysis of the HIV-2 protease, the analysis of two novel protease inhibitors, the creation and analysis of a mutant heterodimeric HIV-2 PR, and the development of a novel expression and purification system for SIV_{AGM} protease which is applicable in principle to all retroviral proteases.

The retroviral proteases are now among the best-characterised enzymes known, and HIV-1 protease in particular has been the target of comprehensive study by a great many groups, both academic and industrial, worldwide. Drugs have been developed to target the HIV proteases and these have proved successful, however, the problem of drug resistant strains has meant that these compounds might ultimately have a limited useful life span. Consequently workers in this field cannot become complacent, and the development of new compounds continues apace. The HIV-1 protease expressed during the course of this project was used to perform preliminary tests on two new inhibitor designs.

Inevitably the urgent requirement for new drugs means that there is a high demand for pure, crystallography grade, retroviral proteases for use in high throughput screens and drug development programs. The yield of these enzymes is typically very poor, as has been discussed, so that production on the scale required for drug discovery is an expensive and difficult process. A means by which these enzymes can be expressed at high levels and purified with ease would consequently have a huge beneficial impact on the drug development process. There would be savings in terms of scale, manpower, time and ultimately costs per unit of enzyme. Academia, the pharmaceutical industry, and their suppliers alike would therefore eagerly embrace a new, high yield expression and purification system for these enzymes. During the course of this work problems of toxicity and poor codon usage have been faced and strategies for overcoming them have been developed. The UDG-SIV_{AGM} protease fusion system, with optimised protease codons, represents possibly the ultimate system for high yield expression and simplified purification of this enzyme. Furthermore the same system could easily be applied to any of the other retroviral

proteases. Consequently this system could be of significant use to workers in the field. The construction of similar vectors for expression of HIV-1 and HIV-2 proteases, each with optimised codons, would be an immediate future aim.

The HIV-2 protease was successfully crystallised at a time when no structural data for this enzyme existed. Unfortunately it was not possible to obtain a structure for the technical reasons already described, and only preliminary data is presented. However, this was certainly one of the earliest (if not the earliest) crystallisations of this enzyme. There are still far fewer structures of HIV-2 protease complexes than there are of HIV-1 protease complexes, and there are no structures of the native HIV-2 protease publically available to date. Paradoxically HIV-2 remains the largest killer between the two HIV serotypes, so that studies specific to HIV-2 protease would potentially provide the greatest benefit to mankind.

The production of mutant HIV-2 PR illustrated the effect of toxicity on expression levels of this enzyme in bacteria. It also demonstrated the effect of a single mutation within the active site on correct folding. It may be possible to design inhibitors able to interact with the enzyme via residues located in the protein's folding core. In this way resistant mutants cannot arise since they would necessarily cause misfolding. A series of site directed mutagenesis experiments could be carried out to determine residues crucial to enzyme folding, and establish whether other residues can be tolerated at these positions. Any positions that cannot tolerate mutation could then be made targets for drug design. The principle of this strategy is the anticipation of resistant mutants prior to drug design in order to eliminate the possiblility that resistance can arise later.

As discussed, early protease inhibitors allowed breakthroughs to be made in the understanding and treatment of the disease (General Introduction, AIDS is a global health problem, page 20 onwards). In the period since 1997 (when work on this project ceased) the use of protease inhibitors has continued to have an impact on our understanding of the disease, and has revolutionised treatment of it. The combined use of RT and PR inhibitors simultaneously has had a dramatic effect. AIDS deaths have been dropping steadily (from their peak in 1995) since combination therapy was introduced (1995 – 1996) with a drop in death rate of 47% in 1997 alone.

Combinatorial therapy eliminated 99% of plasma-borne virus within two weeks (Perelson, Essunger *et al.* 1997), however, this was followed by a much slower phase of decline which suggested that perhaps 3 years of treatment would be required to destroy surviving reservoirs of virus, and many more would be required to eliminate the virus entirely. This was thought to be due to reservoirs of infected cells within the lymphoid tissues, which considerably exceeded the numbers of free and cell-associated virus circulating in the bloodstream. These were shown to be relatively unscathed by existing monotherapies targeting Reverse Transcriptase only (Haase, Henry *et al.* 1996; Cohen, Pantaleo *et al.* 1996). However, a three drug treatment simultaneously targeting both Reverse transcriptase and the viral Protease cleared 99.9% of virus from the lymphoid tissue reservoir within 6 months, eliminating this as the source of the slow second-phase decay of plasma viraemia described by Perelson, Essunger *et al.* It was not clear from this study whether such therapy could completely eliminate the virus, or whether lifelong treatment would have to be maintained, since an alternative reservoir existed. Replication competent, integrated, yet latent, provirus in "resting" (ie. non-dividing) $CD4^+$ cells provides a steady source of new infectious particles as they became activated by host cell division. A coincident study by Chun *et al.* (Chun, Carruth *et al.* 1997) measured this reservoir of latent virus and found that the most prevalent HIV DNA in $CD4^+$ cells was a full length, linear, unintegrated form that was not replication competent. They found a very low load of replication competent, fully integrated, latent provirus among infected $CD4^+$ cells, yet this small reservoir was sufficient to ensure disease progression (Chun, Carruth *et al.* 1997). These results suggested that cessation of combinatorial therapy at any point would result in reinfection, since viral DNA could linger in host cells long after viral load (as measured in terms of RNA) has been eliminated. Less than 0.05% of the resting CD4+ cell population harboured proviral DNA, yet when stimulated to divide many of these were capable of producing infectious virus (Chun, Carruth *et al.* 1997). Furthermore it was not clear how long these cells would persist, it was conceivable that they could survive for months or even years, allowing the virus to evade detection by the immune system and destruction by therapeutic agents.

The combined use of compounds (usually three at once) targeting both RT and PR has since been termed Highly Active Antiretroviral Therapy (HAART), and is currently the best available treatment regime. This regime was shown to largely restore immune

function by allowing the regeneration of $CD4^+$ and CD8 cell levels, although a complete recovery to normal levels was not observed (Autran, Carcelain *et al.* 1997). It has been shown that there is no emergence of drug resistant strains in patients treated with the HAART regime, even after two years (Wong, Hezareh *et al.* 1997; Finzi, Hermankova *et al.* 1997). However, in these same patients there is still a reservoir of infected memory T-cells (<16 cells per million) that when stimulated (using immobilised antibody to CD3 and CD28 – see Spina, Prince *et al.* 1997) produce replicating virus (Wong, Hezareh *et al.* 1997). The observations of Wong *et al.* and Finzi *et al.* support the notion that the slow-phase decay described by Perelson *et al.* is due to the long term survival of latently infected memory T-cells, rather than to a failure of the drug treatment.

To determine whether the reservoir of infected cells would be sufficient to reinitiate infection studies were carried out on patients having already been treated with HAART for over 30 months, and whose plasma viral RNA were observed to be negligible (Chun, Davey *et al.* 1999). On cessation of HAART treatment their plasma viral RNA was seen to increase, demonstrating a re-emergence of infection, within two weeks. Once HAART treatment was resumed the viral RNA levels dropped once more, and had significantly reduced within 2-4 weeks – although they did not reach their former low levels within the duration of the study. Administration of interleukin 2 (Il-2) at intervals during HAART treatment was also studied (Chun, Engel *et al.*, Chun and Fauci 1999). This was intended to stimulate infected memory T-cells, thus forcing integrated provirus out of latency making it vulnerable to the HAART. Measurements showed a decline in the degree of latent infection within the memory T-cell population following the IL-2 treatment, providing a means to reduce the viral reservoir at an increased rate (Chun, Engel *et al.* 1999, Chun, Davey *et al.* Nature brief 1999).

A comprehensive review of this area of AIDS / HIV research up to and including 1999 is provided by Hasse (Haase 1999). Overall it seems that HAART is now an extremely effective strategy for the control of viral infection, and that patients receiving treatment even in the latest stages of AIDS can benefit greatly from treatment and will recover partial immune function (although slowly). However, until means can be developed to eliminate all vestiges of the viral reservoir current therapies simply contain the disease rather than curing it. The development of

Integrase inhibitors may have the desired effect, however, in the meantime something like 5 million people per year are newly infected with HIV worldwide making prevention of infection imperative. Education and welfare efforts are being conducted to this end, but ultimately success may depend on the development of a vaccine – progress towards this goal is reviewed in Ho and Huang 2002.

Appendix

List of suppliers.

Accelrys Ltd. 334 Cambridge Science Park,
Cambridge, CB40WN UK. Phone 01223 228500

Amersham Biosciences (GE Healthcare) (formerly Amersham Pharmacia Biotech UK Ltd)
Amersham Place, Little Chalfont, Bucks. HP7 9NA England. Phone 0870 606 1921
http://www.apbiotech.com

Amicon Amicon is affiliated to Millipore – see Millipore.

Bachem (UK) Ltd. 69 High Street, Saffron Walden, Essex, CB10 1AA, England. Phone +44 / (0) 1799 526465
http://www.bachem.com

BioRad (UK) Bio-Rad United Kingdom Customer Service
Life Science Research
Freephone 0800 181134
Clinical Diagnostics
Freephone 0800 181134
Informatics | Sadtler
Freephone 0800 78945000
http://www.bio-rad.com

Biosoft 2D Dolphin Way, Stapleford, Cambridge, CB2 5DW
http://www.biosoft.com

Cambridge Biosciences Ltd. 5 Signet Court, Swann Rd. Stourbridge Common Business Centre, Cambridge.
Phone: 01223 316855
Fax: 01223 360732

Clontech Labs, Inc. Unit 2, Intec 2, Wade Road, Basingstoke,
Hants. RG24 8NE, UK. Phone 01256 476500
http://www.clontech.com

Costar Distributors in the UK are:
Cole-Parmer Instrument Company Ltd.
Unit 3, River Brent Business Park, Trumpers Way
Hanwell, London W7 2QA
Phone: 0208 574 7556
Fax: 0208 574 7543

Fermentas http://www.fermentas.com
Local distributer is Helena Biosciences Ltd.

Helena Biosciences Ltd. Colina Ave., Sunderland Enterprise Park, Tyne and Wear, SR5 3XB, UK. Phone 0191 549 6054
http://www.helenabiosciences.com

Hellma Hellma GmbH and Co KG
Klosterrunsstrasse 5
D-79371 Müllheim
Phone +49-7631-182-0
Fax +49-7631-13546
Email **info@hellma-worldwide.de**
http://www.hellma-worldwide.com

Inceltech New Brunswick Scientific Co. (below) has acquired the assets and business of Inceltech.

Perkin Elmer Chalfont Road, Seer Green, Beaconsfield,
Bucks. HP9 2FX, UK. Phone 01494 874515
http://uk.instruments.perkinelmer.com

PE-Applied Biosystems UK See Perkin Elmer.

Promega UK Delta House, Chilworth Science Park,
Southampton, SO16 7NS.
Tel: (44) 23 8076 0225
Fax: (44) 23 8076 7014
FREEPHONE: 0800 378994
E-mail Address: ukcustserve@promega.com
Web Address: www.promega.com/uk/

Millipore Millipore (U.K.) Limited,
(Sales and Customer Support Facilities),
Units 3 and 5 The Courtyards, Hatters Lane
Watford, WD18 8YH, UK.
Phone: (44) (19) 238-16375
http://www.millipore.com

New Brunswick 17 Alban Park, Hatfield Road,
St Albans, Herts. AL4 0JJ
Phone: 01727 853855
Fax: 01727 835666
http://www.nbsc.com/

New England Biolabs (UK) Ltd 67 Knowl Piece, Wilbury Way, Hitchin, Herts.
SG4 OTY, UK. Phone 01462 420616
info@uk.neb.com

Sartorius Sartorius Ltd., Longmead Business Centre, Blenheim Rd., Epsom, Surrey, KT19 9QQ. Phone 01372 737100
http://www.sartorius.com/

Shimadzu	European Headquarters Shimadzu Deutschland GmbH Albert-Hahn-Str. 6-10, 47269 Duisburg Tel: +49-203-7687-0 Fax: +49-203-766625 UK Office in Milton Keynes. http://eu.shimadzu.de/home/
Sigma – Aldrich Co. Ltd.	Fancy Rd., Poole, Dorset, BH12 4QH, England. Phone 01202 733114 http://www.sigma-aldrich.com
Stratagene	Stratagene Europe PO Box 12085, 1100 AB Amsterdam, Netherlands. UK customer services 0800 917 3282 http://www.stratagene.com/
QIAGEN UK Ltd.	Boundary Court, Gatwick Rd., Crawley, West Sussex, RH10 9AX. Phone 01293 422 911 (orders) 999 (tech) http://www.qiagen.com/

General Introduction - supplementary information.

The interaction of HIV envelope glycoproteins with CD4 and Chemokine receptors facilitates the viral entry process.

HIV and related retroviruses belong to a class of enveloped fusogenic viruses that include corona-, paramyxo- and orthomyxoviruses (e.g. influenza virus). It is known that the gp41 equivalent (envelope transmembrane) proteins of these viruses are similar in sequence, especially at their N-terminal fusion peptides, and that they are directly involved in membrane fusion. The gp41 ectodomain has the ability to form a coiled coil resembling that of influenza haemagglutinin HA2 (Kowalski, Potz *et al.* 1987, Lu, Blackow *et al.* 1995, Bullough, Hughson *et al.* 1994), supporting the idea that this class of viruses may share some common features of virus entry. However, it is also known that the various enveloped viruses use differing modes of entry; for example direct membrane penetration for HIV, and endocytosis for influenza virus. Even closely related viruses may use unique receptors, consequently the gp120 equivalent (exterior envelope) proteins are also specialised. For this reason there is no similarity in sequence, or structure, between the receptor binding portion of HIV and that of murine leukaemia retrovirus (Fass, Davey *et al.* 1997) for example. Mechanisms for receptor-mediated triggering of fusion may also be virus-specific.

The sequential interaction of the HIV gp120, with CD4 and a chemokine receptor on the host cell surface initiates fusion of the viral and cellular membranes. Entry of a primate immunodeficiency virus into the host cell involves binding of the gp120 envelope glycoprotein to the CD4 glycoprotein, which serves as the primary receptor. The role of gp120 in the process of fusion and entry is crucial. It locates the host cell, anchors the virus to the surface, and orientates the viral spike next to the target membrane. In addition gp120 retains the trimeric gp41 in a metastable conformation and triggers the coordinated release of the three gp41 N-terminal fusion peptides. Elucidation of the gp120 structure is therefore important for an understanding of HIV infection and for the design of therapeutic and prophylactic strategies.

The CD4 receptor has four immunoglobulin-like domains, and structures of the two N-terminal domains (Ryu, Kwong *et al.* 1990, Wang, Yan *et al.* 1990) and the entire

extracellular portion of CD4 (Wu, Kwong and Hendirckson 1997) have been determined. It is the most N-terminal of the four Ig domains that is bound by gp120. Mutagenesis of the CD4 structure analogous to the second Complementarity-Determining Region (CDR2) of immunoglobulins has shown that it is critical for gp120 binding (Moebius, Clayton *et al.* 1992; Sweet, Truneh *et al.* 1991). Conserved gp120 residues important for CD4 binding have also been identified by mutagenesis (Kowalski, Potz *et al.* 1987; Olshevsky, Helseth *et al.* 1990; Cordonnier, Montagnier and Emerman 1989). CD4 binding induces conformational changes in gp120, some either exposing or forming a binding site for specific chemokine receptors. These chemokine receptors, mainly CCR5 and CXCR4 for HIV-1, serve as obligate second receptors for virus entry (Moore 1997; Feng, Broder *et al.* 1996). The gp120 third variable (V3) loop is the principal determinant of chemokine receptor specificity (Speck, Wehrly *et al.* 1997). However, other more conserved gp120 structures exposed on interaction with CD4 are apparently also involved in chemokine-receptor binding. This CD4-induced exposure is indicated by the enhanced binding of several gp120 antibodies (Thali, Moore, *et al.* 1993; Sattentau, Moore *et al.* 1993) which, like V3-loop antibodies, efficiently block the binding of gp120–CD4 complexes to the chemokine receptor (Wu, Gerard, *et al.* 1996). These are called CD4-induced (CD4i) antibodies. CD4 binding may trigger additional conformational changes in the envelope glycoproteins. For example, in some HIV-1 isolates the release of gp120 from the complex is induced (Moore, McKeating *et al.* 1990) - although it is not clear why.

The structure of an intermediate state in the fusion process, during which gp120 is bound to CD4, and that of a final 'fusion-active' state of the gp41 ectodomain have been solved (Kwong, Wyatt *et al.* 1998; Chan, Fass *et al.* 1997; Weissenhorn, Dessen *et al.* 1997). Kwong *et al.* reported the crystal structure at 2.5 Å resolution of a partially deglycosylated HIV-1 gp120 core bound to a two-domain fragment of the CD4 cellular receptor and to the antigen-binding fragment (Fab) of an antibody, 17b, that was directed against a CD4i epitope. The neutralizing antibody, 17b, blocked chemokine-receptor binding. Their accompanying letter related this structure to the antigenic properties of gp120 envelope proteins (Wyatt, Kwong *et al.* 1998). The structure revealed a cavity-laden CD4–gp120 interface, a conserved binding site for

the chemokine receptor, evidence of conformational change upon CD4 binding, the nature of a CD4-induced antibody epitope, and specific mechanisms for immune evasion. Their results provided an insight into the mechanism of HIV entry into host cells.

CD4 binding orients the gp120 surface implicated in chemokine receptor binding to face the target cell, and also forms and exposes the site itself – possibly in a single, concerted shift in the relative orientation of the inner and outer domains. This conformational shift may alter the orientation of the N and C termini, at the proximal end of the inner domain, perhaps partially destabilizing the oligomeric gp120 / gp41 interface (Moore, McKeating *et al.* 1990). Such a shift would also alter the relative placement of the V1/V2 stem (in the CD4i site) and the V3 loop, which emanate from the inner and outer domains respectively. The next step in HIV-1 entry is the interaction of the gp120–CD4 complex with the chemokine receptor. Although interactions between CD4 and chemokine receptor may occur, it is most likely that direct gp120 contacts dominate interaction with the chemokine receptor. As most of the chemokine receptor is encased in the host membrane, binding moves the gp120 bridging sheet closer to the target membrane, a process requiring some CD4 flexibility (Wu, Kwong *et al.* 1997; Moir, Perreault *et al.* 1996). Chemokine-receptor binding is believed to trigger additional conformational changes in the HIV-1 envelope glycoprotein trimer that lead to exposure of the gp41 ectodomain.

Details of the apo state of core gp120, the oligomeric structure, the interaction with the chemokine receptor, the conformational changes that trigger reorganization of the gp41 ectodomain, and the structural basis for insertion of the fusion peptide of gp41 into the target membrane have yet to be determined.

Although gp120 can elicit virus-neutralizing antibodies, HIV eludes the immune system. The highly conserved regions of gp120 create the structures necessary for interaction with the gp41 ectodomain and with receptors on the host cell. In addition five variable regions (V1–V5) have been identified within gp120 (Starcich, Hahn *et*

al. 1986). The first four form surface-exposed loops having disulphide bonds at their bases (Leonard, Spellman, *et al.* 1990). All regions of gp120 are extensively glycosylated (Leonard, Spellman, *et al.* 1990), and the combination of variable regions and overall glycosylation help minimise the immunogenicity and antigenicity of gp120 (Profy, Salinas, *et al.* 1990). Analysis of the antigenic structure of gp120 shows that most of the envelope protein surface is hidden from the immune system by glycosylation and oligomeric occlusion (Wyatt, Kwong, *et al.* 1998). Most broadly neutralizing antibodies access only two surfaces: one that overlaps the CD4-binding site (shielded by the V1/V2 loop), and another that overlaps the chemokine-receptor-binding site (shielded by the V2 and V3 loops). Conformational changes in the gp120 core provide additional mechanisms for evasion from immune surveillance. In the case of the CD4-binding surface, which contains a high proportion of main-chain atoms in the complex, the conformation without CD4 bound may expose underlying side-chain variability. Escape may also be provided by the recessed nature of the binding pocket (steric occlusion) and by a topographical surface mismatch, which encloses a variational island or 'antihotspot'. Mechanisms may be similar in the chemokine-receptor region: conformational change may hide the conserved epitope (unformed before CD4 binding); steric occlusion may take place between the CD4 anchored viral spike and the proximal target membrane; and an 'anti-hotspot' equivalent may camouflage chemokine-receptor-binding residues on the V3 loop in surrounding variability. Some of the means used to elude antibody-based defences may also help HIV avoid cellular immunity. The future design of an effective vaccine against HIV will likely depend on a thorough understanding of the means by which HIV evades the immune system.

The development of Renin inhibitors.

The development of renin inhibitors is in some ways analogous to the development of retroviral protease inhibitors, especially since both targets are aspartic proteases. Consequently it is interesting to consider, briefly, the renin field. Renin catalyses the first and rate limiting step in the renin-angiotensin cascade, ultimately leading to the production of angiotensin II (AII). It cleaves a single peptide bond in the plasma glycoprotein angiotensinogen, and in doing so releases the biologically inactive decapeptide angiotensin I (AI or ANGI). This in turn is the substrate for a second enzyme, Angiotensin Converting Enzyme (ACE), which releases the octapeptide angiotensin II (AII or ANGII). AII is involved in regulation of blood pressure. It is a strong vasoconstrictor and also indirectly stimulates the adrenal cortex, resulting in sodium and water retention. AII's role implicates it in the pathology of hypertension – a major risk factor for cardiovascular diseases such as stroke, myocardial infarction and heart failure. Cardiovascular diseases are the major cause of death in the Western world; consequently the entire renin-angiotensin cascade has been the subject of much research. Renin's position in the cascade (at the first, and rate limiting, step) combined with its unique substrate specificity (it cleaves only one known substrate – angiotensinogen) makes it an ideal target for therapeutic intervention. The renin-angiotensin system and its role in heart failure is explored by the reviews of Wollert and Drexler, and De Mello (Wollert and Drexler 1999; De Mello 2004).

Early renin inhibitors were based on prorenin sequences, pepstatin, or pepstatin analogues. Substrate analogues and inhibitors based on the enzyme-substrate transition state yielded more potent selective inhibitors of renin. Studies in humans demonstrated that such compounds were effective when administered intravenously. However, all the early compounds exhibited poor bioavailability, and rapid elimination when administered orally. There are many reviews of the early developments of this field (Antonaccio and Cushman 1981; Antonaccio 1982; Geiger 1984; Luther, Stein, *et al.* 1989; Kleinert 1989; Wood, Cumin et al. 1994; Lin and Frishman 1996). Improvements in the pharmacokinetic properties of renin inhibitors have recently been achieved by employing the structure based rational design techniques used to great effect in the retroviral protease field. The development of a new class of low molecular weight inhibitor, that does not incorporate the peptide backbone featured in previous inhibitors, was achieved using computer modelling and

crystallographic structure analyses. The first compound in this class is Aliskiren, which has been shown to inhibit AI and AII production in both sodium depleted marmosets and healthy human volunteers ((Nussberger, Wuerzner et al. 2002; Wood, Maibaum, et al. 2003). Subsequent clinical trials on patients with moderate hypertension have demonstrated that it is safe, orally active and effective at lowering blood pressure (Stanton 2003 (1); Stanton 2003 (2); Stanton, Jensen *et al.* 2003).

Neutron Diffraction

Neutrons are neutral particles each with spin 1/2, a magnetic moment of 1.9132 nuclear magnetons, and a half-life of approximately 888 s. Having a magnetic moment causes them to be sensitive to magnetic ordering within a solid. Relative to X-ray diffraction, Neutron diffraction is a fairly new technique - the first neutron diffraction analysis of a protein structure was completed by Schoenborn in 1969 on myoglobin (Schoenborn 1969), and it is still quite uncommon. This is largely because it requires the generation of high thermal-neutron fluxes (a thermal neutron is defined as a neutron possessing a kinetic energy of about 0.025 eV). These can only be obtained from steady state nuclear reactors or from particle accelerators called spallation sources.

Steady State Reactors

A steady-state reactor is essentially a nuclear reactor with a core specially adapted to produce a maximum flux of neutrons. The neutrons are produced by the nuclear fission of ^{235}U. In this process more neutrons are released than are needed to maintain the reaction. The excess neutrons are very high energy "hot or fast neutrons" and must be cooled to generate the "thermal neutrons" suitable for diffraction. The cooling process takes place in a D_2O or graphite moderator kept at a temperature of 40°C. A monochromator is required to select the desired wavelength, resulting in a large loss of neutron flux reaching the sample. The monochromator also provides a beam stop for fast neutrons and gamma-radiation, while a collimator reduces the beam divergence.

Spallation Sources

A Spallation source produces pulsed radiation (typically 25 - 50 Hz). When pulses of high energy protons strike a tungsten or uranium target the result is an emission of neutrons in the order of 20 to 25 per proton. The pulses are very short (~ 4 x 10-16 ns). This allows a high flux and easy cooling. Neutron wavelength is selected by time-of-flight methods, by collecting data at a fixed Bragg angle as a function of neutron energy.

The role of Neutron Diffraction in Structural Biology.

Although X-ray diffraction has certain advantages over neutron diffraction - one being the much higher intensity of the sources available (the greater intensity of X-ray sources allows the use of smaller crystals and facilitates solution of the phase problem) - the principle distinction between X-ray and Neutron diffraction involves the location of hydrogen atoms. Since X-rays are scattered by the atomic electron cloud, atoms with fewer electrons scatter X-rays less effectively. Hydrogen atoms can therefore only be detected by high precision experiments – impossible in the case of macromolecular crystallography. It is also difficult to distinguish neighbouring atoms of the same row of the periodic table and impossible to distinguish between different isotopes of the same element.

Hydrogen represents half of all the atoms in a typical protein, yet a survey of the PDB revealed that only one percent of proteins crystallise well enough to provide sufficient resolution to detect hydrogen atoms using synchrotron X-ray beams.

The greatest advantage of neutron diffraction over X-ray diffraction is its ability to establish the exact position of hydrogen atoms. Since neutron beams interact more strongly with nuclei than do x-rays, neutron diffraction is more useful than x-ray diffraction for determining proton positions. Neutron diffraction can provide data concerning the hydrogen bonds or protonation states of histidine and acidic residues – as demonstrated with lysozyme (Mason, Bentley and McIntyre 1984), the conformations of methyl groups, the solvent structure of the protein, and the degree of protection of particular hydrogen atoms within the structure. An accurate analysis of hydrogen bonding is importance for a complete understanding a protein's structure and activity, and erroneous hydrogen bonding schemes can result from X-ray data due to incorrect assignment of histidine or amide group orientation through an inability to detect hydrogen. Therefore neutron diffraction's ability to directly determine hydrogen / deuterium positions is of immense value for protein structure determination. However, determination of the positions for twice as many atoms - usually with less observed data - than in the case of X-ray diffraction creates additional experimental complexity.

Since neutrons interact with a solid to a much lesser degree than X-rays they therefore possess advantages when studying materials that are damaged by X-rays (such as protein crystals) and in cases where a large penetration depth is desired. Exposure to an X-ray beam leads to radiation induced crystal decay, resulting from the generation of free radicals and their diffusion through the crystal lattice (Blundell and Johnson 1976). Consequently either several crystals, or cryogenic techniques, may be necessary for X-ray protein structure determination. Neutrons, however, do not damage protein crystals so that complete data sets can be collected from a single crystal, and without the need to freeze it. The use of a single crystal is desirable because it eliminates the possibility of scaling errors that might lead to the introduction of artefacts in electron density maps refined using data collected from multiple crystals. It is also easier to grow a single good crystal, than multiple good crystals. However, since the bulk of neutrons in a beam will pass through a sample without interaction, neutron diffraction requires high flux and large crystals in order to produce sufficient data. A larger crystal provides a greater probability of interaction between the atoms within the crystal and the neutrons passing through it.

Crystal growth remains the bottleneck for protein structure determination by any diffraction method, but for neutron diffraction the problem is exacerbated by the requirement for much larger crystals than those suitable for X-ray techniques. Recent advancements in neutron diffraction methods and instrumentation are beginning to allow the use of much smaller crystals or the determination of much larger unit cells. While the early neutron diffraction experiments used crystals up to 30 mm^3, single crystals with a volume of 1 mm^3 can now be studied. Since the inception of protein crystallisation experiments in space on board the U.S. Space Shuttle Columbia in 1995 and on the Russian Mir space station in the late 1990s, the growth of larger crystals in low gravity has become more common. Proposals have been submitted to NASA to grow large crystals for neutron studies in the International Space Station. Techniques for growing crystals under conditions of microgravity on earth are also being developed. For most proteins though, it is still difficult to obtain crystals large enough for neutron diffraction measurement.

An additional complication is that in comparison with X-ray diffraction there is an increased contribution of solvent scattering effects, and overall scattering is much higher for neutron diffraction than for X-ray diffraction. For hydrogen atoms most of the scattering is incoherent and thus does not contribute to the reflection intensity but only to the background, and is the major source of background for neutron diffraction. A solution involves replacement of hydrogen atoms by deuterium atoms to reduce the background noise and simultaneously increase the coherent scattering power (i.e. that which contributes to reflection intensity) in neutron diffraction. For this reason the substitution of hydrogen with deuterium by soaking crystals in D_2O for long periods (often many months), and more recently the growth of perdeuterated crystals (Gamble, Clauser and Kossiakoff, 1994) has been necessary to reduce background. In order to obtain fully perdeuterated protein, large quantities are expressed in *E. coli* grown using minimal medium containing deuterated amino acids and deuterated water. Crystallisation conditions, space group and cell parameters have been demonstrated to be the same for both native and perdeuterated forms of a protein and no significant structural differences between the forms are observed.

X-ray and neutron diffraction techniques are in many senses complementary and used together they represent the most effective direct method for the determination of protein structures. Consequently neutron diffraction studies are usually carried out in conjunction with X-ray diffraction experiments. The X-ray model can be used as a starting point for a separate refinement of the neutron data - either with or without the calculated hydrogen positions (Kossiakoff and Spencer 1981; Phillips and Schoenborn 1981), or the two data sets can be used for a simultaneous joint refinement (Wlodawer, Walter, *et al.* 1984).

At present very few protein structures have been solved using neutron diffraction. However, with more neutron diffraction facilities becoming available worldwide that could soon change. One such new facility, the Spallation Neutron Source (SNS), is due to be completed in 2006. It will be ten times brighter than existing neutron facilities and ideal for atomic-resolution protein crystallography studies. Neutron diffraction studies of protein structures are also being facilitated by the availability of improved neutron detectors that speed up data collection and by improvements in the ability to grow larger crystals. It is estimated that within a decade the number of protein structures determined annually using neutron diffraction could reach 50-100, although this is still a very small proportion of the structures solved by X-ray diffraction. It is likely that due to the larger experimental cost and effort required, neutron diffraction will remain limited in its use, and will predominantly be applied to the resolution of specific issues, where X-ray diffraction is inappropriate.

Low Barrier Hydrogen Bonds

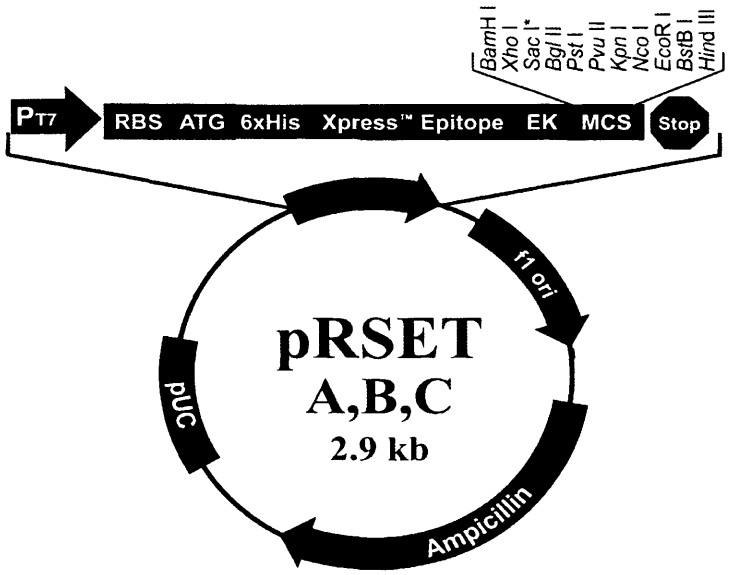
The concept of a LBHB was introduced by Cleland following observation of low hydrogen fractionation factors for certain enzymes. He suggested that the potential energy surface of some hydrogen bonds is a double well potential with a low barrier (~2 kcal/mol) toward hydrogen transfer, which he termed a Low-Barrier Hydrogen Bond (Cleland 1992). Gerlt and Gassman (Gerlt and Gassman 1993, Biochemistry) also suggested the presence of special hydrogen bonds, proposing a scenario in which the hydrogen resides midway between the proton acceptor and donor. Cleland and Kreevoy suggested that LBHBs could prove to be very strong, with bonding energy as high as 10-20 kcal/mol, effectively making them covalent bonds (Cleland and Krevoy 1994), hence the alternative name Short-Strong Hydrogen Bond (SSHB). Investigations of the presence and properties of LBHBs have ensued. In general they are characterised by great strength, short length, and potential surfaces having a very small (if any) barrier to proton migration. It is proposed that they are formed from a conventional hydrogen bond when the proton donor is positively charged, or the acceptor is negatively charged, (hence they may also be called positive-ion or negative-ion hydrogen bonds). When hydrogen bonds are as short as 2.5 – 2.6 Å the proximity of the donor and acceptor atoms reduces the energy barrier which normally prevents transfer of the hydrogen from the donor to the acceptor group. Theoretical calculations on the system (O- - - H+- - -O) suggest that the potential energy barrier becomes less than 5 kcal mol-1 when the O- - -O distance is less than 2.8Å, thus permitting easy proton transfer. This allows rapid exchange of the proton between donor and acceptor to occur, and it is this proton mobility that may be of significance mechanistically. LBHBs are of particular interest to biologists because it has been hypothesised that they are involved in enzyme catalysis (Cleland 1992; Cleeland and Kreevoy 1994; Frey, Whitt and Tobin 1994), and that their formation may stabilize the intermediates and / or transition states of enzymatic reactions (Cleland 1992; Gerlt and Gassman 1993; Cleland and Krevoy 1994; Frey, Whitt and Tobin 1994; Tobin, Whitt *et al.* 1995; Cassidy, Lin and Frey 1997; Zhao, Abeygunawardana *et al.* 1996; Zhao, Abeygunawardana *et al.* 1997; Tong and Davis 1995; Hur, Leja and Dunn 1996; Kahyaoglu, Haghjoo *et al.* 1997) explaining why an enzyme is able to bind intermediates or transition states more tightly than the original substrate. While there is increasing evidence to support this idea, it has not been without controversy (Scheiner 1992; Warshel, Papazyan and Kollman 1995; Warshel and Papazyan 1996).

Materials and methods – supplementary information.

pRSET vector system for the expression of recombinant proteins.

The pRSET system (fig. 1) from Invitrogen (www.invitrogen.com) is designed to allow high level expression of genes in *E.coli* and subsequent purification of expressed protein by IMAC. The pUC derived vectors contain a strong T7 phage promoter and a multiple cloning site which allows genes to be inserted down stream of and in frame with a sequence encoding an N-terminal fusion peptide. This fusion consists of a polyhistidine tag (allowing IMAC), a transcript stabilising sequence from gene 10 of page T7, an Anti-Xpress epitope, and an enterokinase cleavage recognition sequence (which allows later enzymatic removal of the N-terminal fusion peptide). Three versions of the pRSET vector are available (designated A, B, and C) which differ only in a single variable region designed to allow cloning into each of the three possible reading frames whilst maintaining the correct fusion sequence. Version B was used throughout this project. The T7 promoter is recognised by T7 RNA polymerase, which must be expressed by the host strain (either via IPTG induction of the polymerase gene carried on the DE3 bacteriophage lambda lysogen incorporated in some strains, or as a result of deliberate infection by phage expressing the polymerase). Strains incorporating the DE3 lysogen were used for any work involving this system described herein (specifically BL21(DE3) unless otherwise stated). Such strains always exhibit a basal level of T7 RNA polymerase expression, which can result in premature expression of the target protein. T7 lysozyme is able to bind T7 polymerase and inhibit transcription, thus the incorporation of the pLysS plasmid (confering chloramphenicol resistance), which encodes T7 lysozyme, can be used to minimise unwanted polymerase activity prior to induction. Additionally the T7 lysozyme is able to cleave the peptidoglycan layer of the *E.coli* cell wall and thus aids cell lysis, especially when used in conjunction with freeze thawing of harvested cells.

Figure 1 The pRset vector map and polylinker region sequence.



*Version C does not contain *Sac* I

Comments for pRSET A
2897 nucleotides

T7 promoter: bases 20-39
6x His tag: bases 112-129
T7 gene 10 leader: bases 133-162
Anti-Xpress™ epitope: bases 169-192
Multiple cloning site: bases 202-248
pRSET reverse priming site: bases 295-314
T7 transcription terminator: bases 256-385
f1 origin: bases 456-911
bla promoter: bases 943-1047
Ampicillin (*bla*) resistance gene (ORF): bases 1042-1902
pUC origin (C): bases 916-2852
(C) = Complementary strand

T7 promoter — RBS

```
AATACGACTC ACTATAGGGA GACCACAACG GTTTCCCTCT AGAAATAATT TTGTTTAACT TTAAGAAGGA
```

Polyhistidine (6xHis) region

```
GATATACAT ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT
          Met Arg Gly Ser His His His His His His Gly Met Ala Ser  Met Thr
```

T7 gene 10 leader — Xpress™ Epitope — *Bam*H I — *Xho* I *Sac* I

```
GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCG AGC TCG
Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr Asp Asp Asp Asp Lys Asp Pro Ser Ser
```

EK recognition site — EK cleavage site

Bgl II — *Pst* I — *Pvu* II — *Kpn* I *Nco* I — *Eco R* I — *Bst*B I *Hind* III

```
AGA TCT GCA GCT GGT ACC ATG GAA TTC GAA GCT TGA TCCGGCTGCT AACAAAGCCC
Arg Ser Ala Ala Gly Thr Met Glu Phe Glu Ala ***
```

pRSET reverse priming site

```
GAAAGGAAGC TGAGTTGGCT GCTGCCACCG CTGAGCAATA ACTAGCATAA
```

Results chapters – supplementary information.

SPA assay data.

Explanation of appendix tables 1 to 5.

The tables 1-5 show raw data for SPA assays carried out on fractions collected during purification of HIV-1 and HIV-2 proteases, as described in the text of chapters 3 and 4. Tables 1 and 2 relate to the elution of HIV-1 PR from an S-200 Size exclusion column, table 3 to the elution of HIV-2 PR from the same column. Table 4 refers to the elution of HIV-2 PR from a Phenyl Sepharose column, and table 5 to the elution of HIV-2 PR from an S-Sepharose column. The following abbreviations / symbols were used in these tables.

Bg Background counts, data collected from empty vials.

0% Control samples containing buffer and protein but no SPA reagent. Such samples produce no scintillation, and any counts are the result of background

100% The maximum scintillation counts possible, determined by counting vials containing buffer and SPA reagent, but no protease sample.

Rep. Repetition (ie. a duplicate sample reading).

Samples were represented by their fraction numbers and were counted in duplicate (reps 1- 2) and an average taken. Multiple background counts were made and an average taken. 0% and 100% controls were distributed at the start, middle, and end of the run to account for any possible instrumental drift, and an average figure for each was calculated. In table 2 the average background count was subtracted from each of the average sample counts and the result expressed as a percentage of total possible scintillation. The higher the value for "% uncut SPA substrate", the lower the level of activity in the sample. The % relative activity for each sample was obtained by subtracting the figure for % uncut SPA substrate from 100. A high value for this figure indicates a greater proteolytic activity in the fraction. It can therefor be seen that fraction 10 in table 2 contains the most protease activity. In subsequent analyses (tables 3-5) a single count was taken for each sample, at the start and end of the incubation period, and the relative activity was calculated as described here.

Table 1 HIV-1 PR SPA raw data (Chapter 3 Figure 4).

See page 342 for explanation of Appendix tables 1-5, and see also Chapter 3, figure 4, page 192.

Detection of HIV-1 PR activity by SPA assay.

Scintillation counts for fractions eluted from Sephacryl 200 column.

Fraction no. & controls	Rep. 1 Counts per minute (CPM)	Rep. 2 CPM	Averages
0%	25.00	161	93.00
100%	10963.16	10976.84	10970.00
1	10283.00	10196	10239.50
2	9576.00	9551	9563.50
3	11808.24	12228.24	12018.24
4	10464.00	9983	10223.50
5	11016.84	10916.84	10966.84
6	9714.00	9817	9765.50
7	6765.00	6868	6816.50
8	3218.00	3331	3274.50
9	2539.00	2665	2602.00
100%	11323.33	11132.22	11227.78
100%	13560.00	13141.25	13350.63
10	2181.00	2233	2207.00
11	3145.00	3069	3107.00
12	2436.00	2314	2375.00
13	2313.00	2335	2324.00
14	3010.00	3029	3019.50
15	3398.00	3267	3332.50
16	4152.00	4117	4134.50
17	4348.00	4242	4295.00
18	5718.00	5975	5846.50
19	7530.00	8133	7831.50
100%	10951.58	10307	10629.29
100%	10808.42	9692	10250.21
20	8629.00	8508	8568.50
21	8575.00	8683	8629.00
0%	15.00	24	19.50
Background (Bg)	33.00	16.00	24.50
Bg	16.00	86.00	51.00
Bg	45.00	39.00	42.00
Bg	118.00	74.00	96.00
Bg	67.00	72.00	69.50
Bg	20.00	31.00	25.50
Bg	38.00	22.00	30.00
Average Bg	48.14	48.57	48.36
Average 0%	20	92.5	56.25
Average 100%	11521.3	11049.86	11285.58

Table 2 HIV-1 PR SPA relative activity data (Chapter 3 Figure 4).

See page 342 for explanation of Appendix tables 1-5, and see also Chapter 3, figure 4, page 193.

Detection of HIV-1 PR activity by SPA assay.

Calculation of relative protease activity for each fraction.

Fraction	CPM	CPM - Bg	% uncut SPA substrate	% relative activity
1	10239.50	10191.14	90.75	9.25
2	9563.50	9515.14	84.73	15.27
3	12018.24	11969.88	106.59	-6.59
4	10223.50	10175.14	90.61	9.39
5	10966.84	10918.48	97.23	2.77
6	9765.50	9717.14	86.53	13.47
7	6816.50	6768.14	60.27	39.73
8	3274.50	3226.14	28.73	71.27
9	2602.30	2553.94	22.74	77.26
10	2207.00	2158.64	19.22	80.78
11	3107.00	3058.64	27.24	72.76
12	2375.00	2326.64	20.72	79.28
13	2324.00	2275.64	20.27	79.73
14	3019.50	2971.14	26.46	73.54
15	3332.50	3284.14	29.25	70.75
16	4134.50	4086.14	36.39	63.61
17	4295.00	4246.64	37.82	62.18
18	5846.50	5798.14	51.63	48.37
19	7831.50	7783.14	69.31	30.69
20	8568.50	8520.14	75.87	24.13
21	8629.00	8580.64	76.41	23.59
100%	11285.58	11237.22	100	0
0%	56.25	7.89	0	100
1%	112.2933	112.2933	1	99

Where:-

1% = (CPM 100% - CPM 0%) / 100

% relative activity = 100 - % uncut SPA substrate

Table 3 HIV-2 PR SPA raw and relative activity data (Chapter 4 Figure 4).

See page 342 for explaination of Appendix tables 1-5, and see also Chapter 4, Figure 4, page 214.

Detection of HIV-2 PR activity by SPA assay.

Scintillation counts for fractions eluted from Sephacryl 200 column.

Fraction no. & controls	Time zero Counts / min	12 hrs @ 37 C Counts / min	Difference	% activity
1	16043.08	14717.15	1325.93	8.1
2	14795.72	12927.51	1868.21	11.3
3	16415.40	15380.01	1035.39	6.3
4	14848.59	13698.68	1149.91	7.0
5	16392.33	15285.74	1106.59	6.7
6	15091.45	12220.02	2871.43	17.4
7	16253.87	14109.36	2144.51	13.0
8	17443.36	13430.69	4012.67	24.4
9	17811.70	13058.78	4752.92	28.9
10	18574.58	11420.02	7154.56	43.5
11	18632.79	10225.02	8407.77	51.1
12	16828.38	9525.03	7303.35	44.4
13	16524.66	8337.02	8187.64	49.7
14	16441.59	8602.03	7839.56	47.6
15	15720.05	10164.03	5556.02	33.7
16	14892.89	13690.71	1202.18	7.3
17	15002.93	14591.48	411.45	2.5
18	15680.08	15972.37	-292.29	-1.8
19	14607.22	15178.63	-571.41	-3.5

Bg	19.00	14.00	15.00	12.00
0%	22.00	21.00	23	20
100%	15430.83	16358.47	16811.74	17343.34

Avg Bg	15.0	
Avg. 0%	21.5	
Avg. 100%	16486.1	
Maximum difference		16464.6
1% activity		164.65

1% = (CPM 100% - CPM 0%) / 100

% relative activity = 100 - % uncut SPA substrate

Table 4 HIV-2 PR SPA raw and relative activity data (Chapter 4 figure 7).

See explanation of Appendix tables 1-5, page 342; and Chapter 4, fig. 7, page 218.

Detection of HIV-2 PR activity by SPA assay.

Scintillation counts for fractions eluted from Phenyl Sepharose column.

Fraction	Max. CPM	Sample CPM	Difference	% activity
1	8057.79	8542.01	-484.22	-2.9
2	8057.79	7748.01	309.78	1.9
3	8057.79	7208.01	849.78	5.2
4	8057.79	8192.01	-134.22	-0.8
5	8057.79	7099.01	958.78	5.8
6	8057.79	8382.01	-324.22	-2.0
7	8057.79	7499.02	558.77	3.4
8	8057.79	8919.02	-861.23	-5.2
9	8057.79	8590.02	-532.23	-3.2
10	8057.79	8235.02	-177.23	-1.1
11	8057.79	5904.02	2153.77	13.1
12	8057.79	1694.01	6363.78	38.7
13	8057.79	2747.01	5310.78	32.3
14	8057.79	3794.01	4263.78	25.9
15	8057.79	4564.02	3493.77	21.2
16	8057.79	5614.02	2443.77	14.8
17	8057.79	7655.03	402.76	2.4
18	8057.79	7051.03	1006.76	6.1
19	8057.79	7242.03	815.76	5.0
20	8057.79	6911.03	1146.76	7.0
21	8057.79	7229.04	828.75	5.0

Repetitions

Bg	18.00	14.00	16.00	
0%	15.00	14.00	16.00	
100%	7680.00	9185.05	8164.05	7202.04

Avg Bg	16.00
Avg. 0%	15.00
Avg. 100%	8057.79

Maximum difference	8042.785
1% activity	80.43

1% = (CPM 100% - CPM 0%) / 100

% relative activity = 100 - % uncut SPA substrate

Table 5 HIV-2 PR SPA raw and relative activity data (Chapter 4 figure 9).

See explanation of Appendix tables 1-5 page 342, and Chapter 4, figure 9 page 221.

Peak No.	Fraction	CPM	CPM-BG	% uncut	% relative activity
1	1	7490	7478	101	-1
1	2	7441	7429	100	0
1	3	7646	7634	103	-3
2	4	6636	6624	89	11
2	5	6090	6078	82	18
3	6	2708	2696	36	64
3	7	3085	3073	41	59
3	8	3039	3027	41	59
3	9	2976	2964	40	60
3	10	3492	3480	47	53
3	11	4331	4319	58	42
3	12	5302	5290	71	29
3	13	4796	4784	65	35
3	14	5615	5603	76	24
3	15	4695	4683	63	37
4	16	5132	5120	69	31
4	17	5568	5556	75	25
5	18	5548	5536	75	25
5	19	6357	6345	86	14
6	20	5206	5194	70	30
6	21	4699	4687	63	37
6	22	5383	5371	73	27
7	23	5240	5228	71	29
8	24	6142	6130	83	17
8	25	5467	5455	74	26
8	26	7423	7411	100	0
9	27	7912	7900	107	-7
9	28	6846	6834	92	8
9	29	9517	9505	128	-28
9	30	7006	6994	94	6

Backgd	0%	Max (100%)	1%
12	23	7406	74.06

Chromogenic activity assay data.

Explanation of appendix tables 6 to 8.

The content of tables 6 to 8 (pages 349-351) represents the raw data for chromogenic activity assays carried out on both mutant heterodimer HIV-2 PR and native HIV-2 PR, as described in the text of Chapter 5 and illustrated in figure 23 (page 266) of the same chapter.

Table 6 Analysis of mutant HIV-2 PR for activity (Chapter 5 figure 23).

For explanation see page 348, and see Chapter 5, figure 23, page 266.

Simple Mix activity assay
His / Cys HIV-2 PR mutant heterodimer vs substrate VII

0 M NaCl

pH	Abs 0 hrs	Abs 20 hrs	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.040	0.047	-0.007	-17.5	-16.6
4	0.038	0.044	-0.006	-15.8	
5	0.041	0.044	-0.003	-7.3	-8.4
5	0.042	0.046	-0.004	-9.5	
6	0.042	0.049	-0.007	-16.7	-13.0
6	0.043	0.047	-0.004	-9.3	
7	0.044	0.049	-0.005	-11.4	-9.3
7	0.042	0.045	-0.003	-7.1	
8	0.060	0.069	-0.009	-15.0	-13.1
8	0.054	0.060	-0.006	-11.1	

1 M NaCl

pH	Abs 0 hrs	Abs 20 hrs	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.042	0.046	-0.004	-9.5	-4.8
4	0.046	0.046	0.000	0.0	
5	0.040	0.045	-0.005	-12.5	-9.8
5	0.042	0.045	-0.003	-7.1	
6	0.039	0.040	-0.001	-2.6	-5.8
6	0.044	0.048	-0.004	-9.1	
7	0.046	0.052	-0.006	-13.0	-8.6
7	0.047	0.049	-0.002	-4.3	
8	0.059	0.063	-0.004	-6.8	-9.2
8	0.052	0.058	-0.006	-11.5	

Table 7 Analysis of mutant HIV-2 PR for activity (Chapter 5 figure 23).

For explaination see page 348, and see Chapter 5, figure 23, page 266.

Refolded mixture activity assay
His / Cys HIV-2 PR mutant heterodimer vs substrate VII

0 M NaCl

pH	Abs 0 hrs	Abs 20 hrs	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.059	0.061	-0.002	-3.9	-16.4
4	0.045	0.058	-0.013	-28.9	
5	0.047	0.054	-0.007	-15.8	-13.9
5	0.050	0.056	-0.006	-12.0	
6	0.071	0.078	-0.007	-10.5	-8.2
6	0.068	0.072	-0.004	-5.9	
7	0.077	0.079	-0.002	-2.6	-2.6
7	0.086	0.088	-0.002	-2.6	
8	0.080	0.081	-0.001	-1.3	-5.3
8	0.075	0.082	-0.007	-9.3	

1 M NaCl

pH	Abs 0 hrs	Abs 20 hrs	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.065	0.070	-0.005	-7.7	-6.6
4	0.081	0.085	-0.004	-5.5	
5	0.062	0.076	-0.014	-21.7	-11.7
5	0.058	0.059	-0.001	-1.7	
6	0.044	0.049	-0.005	-10.9	-12.7
6	0.043	0.049	-0.006	-14.6	
7	0.049	0.054	-0.005	-10.0	-7.9
7	0.053	0.056	-0.003	-5.9	
8	0.062	0.062	0.000	-0.2	-4.4
8	0.058	0.063	-0.005	-8.6	

Table 8 Analysis of control HIV-2 PR for activity (Chapter 5 figure 23).

For explanation see page 348, and Chapter 5, figure 23, page 266.

Control activity assay
HIV-2 PR vs substrate VII

0 M NaCl

pH	Abs 0 s	Abs 180 s	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.072	0.068	0.004	5.6	7.6
4	0.073	0.066	0.007	9.6	
5	0.124	0.110	0.014	11.3	8.9
5	0.107	0.100	0.007	6.5	
6	0.097	0.090	0.007	7.2	5.0
6	0.071	0.069	0.002	2.8	
7	0.055	0.054	0.001	1.8	-1.9
7	0.053	0.056	-0.003	-5.7	
8	0.073	0.075	-0.002	-2.7	-4.3
8	0.068	0.072	-0.004	-5.9	

1 M NaCl

pH	Abs 0 s	Abs 180 s	Abs. Diff.	% Rel. Act.	Avg. Rel. Act.
4	0.045	0.037	0.008	17.8	16.3
4	0.047	0.040	0.007	14.9	
5	0.055	0.045	0.010	18.2	16.9
5	0.051	0.043	0.008	15.7	
6	0.089	0.084	0.005	5.6	12.4
6	0.052	0.042	0.010	19.2	
7	0.063	0.063	0.000	0.0	-0.6
7	0.078	0.079	-0.001	-1.2	
8	0.112	0.114	-0.002	-1.8	-1.5
8	0.089	0.090	-0.001	-1.1	

Further Reading

Table 9 Some useful crystallography texts

• Crystallography Made Crystal Clear	(Rhodes 1993)
• Crystal Structure Analysis	(Glusker and Trueblood 1985)
• Practical Protein Crystallography	(McRee 1999)
• Crystallographic Methods And Protocols	(Jones, Mulloy *et al.* 1996)

References

Article: (1987). "AIDS virus cultured [news]." Nature **330**(6145): 197.

Article: (1992). "Special issue: Transposable elements and evolution." Genetica **86** (1-3): 1-311.

Abdel-Meguid, S. S., B. Zhao, *et al.* (1993). "Inhibition of human immunodeficiency virus-1 protease by a C2-symmetric phosphinate. Synthesis and crystallographic analysis." Biochemistry **32**(31): 7972-80.

Akari H., S. Bour, *et al.* (2001). "The human immunodeficiency virus type 1 accessory protein Vpu induces apoptosis by suppressing the nuclear factor kappaB-dependent expression of antiapoptotic factors." J. Exp Med. **194**(9):1299-311.

Ala, P. J., R. J. DeLoskey, *et al.* (1998). "Molecular recognition of cyclic urea HIV-1 protease inhibitors." J Biol Chem **273**(20): 12325-31.

Ala, P. J., E. E. Huston, *et al.* (1997). "Molecular basis of HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors." Biochemistry **36**(7): 1573-80.

Ala, P. J., E. E. Huston, *et al.* (1998). "Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities." Biochemistry **37**(43): 15042-9.

Allan, J.S. (1991). "Pathogenic properties of simina immunodeficiency viruses in nonhuman primates." W. Koff *et al.*, Annual review of AIDS research, New York, Marcel Dekker, **1**:191-206.

Almond, N., A. Jenkins, *et al.* (1992). "Population sequence analysis of a simian immunodeficiency virus(32H reisolate of SIVmac251): a virus stock used for international vaccine studies." AIDS Res. Human Retroviruses **8**: 77-88.

Alonso, A., T. P. Cujec, *et al.* (1994). "Effects of human chromosome 12 on interactions between Tat and TAR of human immunodeficiency virus type 1." J Virol **68**(10): 6505-13.

Antanaccio, M.J., (1982). "Inhibitors of the renin - angiotensin system as new antihypertensive agents." Clin. Exp. Hypertens. A. **4**(1-2):27-46.

Antanaccio, M.J. and D.W. Cushman, (1981). "Drugs inhibiting the renin - angiotensin system." Fed. Proc. **40**(8):2275-84.

Antonov, V. K., L. M. Ginodman, *et al.* (1978). "Mechanism of pepsin catalysis: general base catalysis by the active-site carboxylate ion." FEBS Lett. **88**(1): 87-90.

Antonov, V. K., L. M. Ginodman, *et al.* (1981). "Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange." Eur. J. Biochem. **117**(1): 195-200.

Arnold, G. J., K. Popinga, *et al.* (1994). "Inhibition of HIV-1 proteaseby bifunctional interface peptides." Peptides 1994. H. L. S. Maia. Leiden**:** 345-346.

Arthur, L. O., J. W. Bess, Jr., *et al.* (1992). "Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines." Science **258**(5090): 1935-8.

Ascher, M. S., H. W. Sheppard, *et al.* (1995). "HIV results in the frame. Paradox remains." Nature **375**(6528): 196; author reply 198.

Ashorn, P., T. J. McQuade, *et al.* (1990). "An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection." Proc Natl Acad Sci U S A **87**(19): 7472-6.

Autran, B., G. Carcelain, *et al.* (1997). "Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease." Science **277**(5322): 112-6.

Babé, L. M. and C. S. Craik (1991). "Time dependent heterodimer formation leads to inhibition of HIV protease activity." Adv Exp Med Biol **306**: 543-7.

Babé, L. M., S. Pichuantes, *et al.* (1991). "Inhibition of HIV protease activity by heterodimer formation." Biochemistry **30**(1): 106-11.

Babé, L. M., J. Rose, *et al.* (1992). "Synthetic interface peptides alter dimeric assembly of the HIV 1 and 2 proteases." Protein Sci. **1**: 1244-1253.

Babé, L. M., J. Rose, *et al.* (1995). "Trans-dominant inhibitory human immunodeficiency virus type 1 protease monomers prevent protease activation and virion maturation." Proc Natl Acad Sci U S A **92**(22): 10069-73.

Babine, R.E., N. Zhang, *et al.* (1993). "Structure-activity studies on pseudo-symmetrical HIV-1 Protease Inhibitors." Bioorg. Med. Chem. Lett. **3**: 1589-1594.

Baca, M. and S. B. H. Kent (1993). "Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis." Proc. Natl. Acad. Sci. USA **90**: 11638-11642.

Baca, M. and S. B. H. Kent (2000). "Protein backbone engineering through total chemical synthesis: new insight into the mechanism of HIV-1 protease catalysis." Tetrahedron **56**: 9503-9513.

Bachovchin, W.W. (1985). "Confirmation of the assignment of the low-field proton resonance of serine proteases by using specifically nitrogen-15 labelled enzyme." Proc. Natl. Acad. Sci. USA **82**: 7948-7951.

Back, N. K., M. Nijhuis, *et al.* (1996). "Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme." Embo J **15**(15): 4040-9.

Bagossi, P., Y. S. Cheng, *et al.* (1996). "Activity of linked HIV-1 proteinase dimers containing mutations in the active site region." Protein Eng **9**(11): 997-1003.

Bagossi, P., Y. S. Cheng, *et al.* (1998). "Comparison of the specificity of homo- and heterodimeric linked HIV-1 and HIV-2 proteinase dimers." Prot. Eng. **11**(6): 439-45.

Baker, E. N. and J. Drenth (1987). "The thiol proteases: structure and mechanism. Biological Macromolecules and Assemblies." F. A. Jurnak and A. McPherson. New York, John Wiley. **3:** 313-368.

Baldwin, E. T., T. N. Bhat, *et al.* (1995). "Structure of HIV-1 protease with KNI-272: a transition state mimetic inhibitor containing allophenylnorstatine." Adv Exp Med Biol **362**: 445-9.

Baldwin, E. T., T. N. Bhat, *et al.* (1995). "Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine." Structure **3**(6): 581-90.

Baldwin, E. T., T. N. Bhat, *et al.* (1995). "Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase." Nat Struct Biol **2**(3): 244-9.

Barre-sinoussi, F., J. C. Chermann, *et al.* (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for AIDS." Science **220**: 868-871.

Baum, E. Z., G. A. Bebernitz, *et al.* (1990). "Isolation of mutants of human immunodeficiency virus protease based on the toxicity of the enzyme in *Escherichia coli*." Proc Natl Acad Sci U S A **87**(14): 5573-7.

Bayer, P., M. Kraft, *et al.* (1995). "Structural studies of HIV-1 Tat protein." J Mol Biol **247**(4): 529-35.

Bayliss, R. S., J. R. Knowles, *et al.* (1969). "An aspartic acid residue at the active site of pepsin. The isolation and sequence of the heptapeptide." Biochem. J. **113**(2): 377-86.

Berger, A. and I. Schechter (1970). "Mapping the active site of papain with the aid of peptide substrates and inhibitors." Philos. Trans. R. Soc. Lond. B. Biol. Sci. **257**(813): 249-264.

Beveridge, A. J. and G. C. Heywood (1993). "A quantum mechanical study of the active site of aspartic proteinases." Biochemistry **32**(13): 3325-33.

Bhat, T. N., E. T. Baldwin, *et al.* (1995). "X-ray structure of a tethered dimer for HIV-1 protease." Adv Exp Med Biol **362**: 439-44.

Biberfeld, G., F. Brown, *et al.* (1987). "WHO Working Group on Characterization of HIV-Related Retroviruses: criteria for characterization and proposal for a nomenclature system." Aids **1**(3): 189-90.

Billich, A., F. Hammerscmid, *et al.* (1990). "Purification, assayand kinetic features of HIV-1 proteinase." Biol Chem Hoppe Seyler **371**(3): 265-72.

Bleul, C. C., M. Farzan, *et al.* (1996). "The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry." Nature **382**(6594): 829-33.

Blumenstein, J. J., T. D. Copeland, *et al.* (1989). "Synthetic non-peptide inhibitors of HIV protease." Biochem Biophys Res Commun **163**(2): 980-7.

Blundell, T. L., J. Cooper, *et al.* (1987). "On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogues." Biochemistry **26**(18): 5585-90.

Blundell TL, Johnson LN, (1976). "Protein Crystallography" Academic Press.

Blundell, T. L., H. B. Jones, *et al.* (1980). "Enzyme regulation and mechanism of action." P. Mildner and B. Ries. Oxford, Pergamon: 281-288.

Blundell, T., B. L. Sibanda, *et al.* (1983). "Three-dimensional structure, specificity and catalytic mechanism of renin." Nature **304**(5923): 273-5.

Boggetto, N. and Reboud-Ravaux, M. (2002). "Dimerization inhibitors of HIV-1 protease." Biol. Chem. **383**(9):1321-4.

Bone, R., J.P. Vacca, *et al.* (1991). "X-ray crystal-structure of the HIV protease complex with L-700,417, an inhibitor with pseudo C2 symmetry." J. Am. Chem. **113**: 9382-9384.

Bonekamp, F., and K.F. Jensen (1988). "The AGG codon is translated slowly in *E.coli* even at very low expression levels." Nucleic Acids Res. **16**(7): 3013-24.

Bott, R., E. Subramanian, *et al.* (1982). "Three-dimensional structure of the complex of the Rhizopus chinensis carboxyl proteinase and pepstatin at 2.5-A resolution." Biochemistry **21**(26): 6956-62.

Boucher, C. A. B., E. O'Sullivan, *et al.* (1992). "Ordered appearance of Zidovudine resistance mutations during the treatment of 18 Human Immunodeficiency Virus-positive subjects." Journal of infectious diseases **165**: 105-110.

Bouma, J. E. and R. E. Lenski (1988). "Evolution of a bacteria/plasmid association." Nature **335**(6188): 351-2.

Bouras, A., N. Boggetto, *et al.* (1999). "Design, synthesis, and evaluation of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerization." J Med Chem **42**(6): 957-62.

Bowie, J. U. and R. T. Sauer (1989). "Equilibrium dissociation and unfolding of the Arc repressor dimer." Biochemistry **28**: 7140-7143.

Bowman, M.J. and Chmielewski, J. (2002). "Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides." Biopolymers. **66**(2):126-33.

Brocklehurst, K., B. S. Baines, *et al.* (1981). "Papain and other constituents of carica papaya L." Top. Enzyme. Fermentation Biotechnol. **5**: 262-335.

Brocklehurst, K. and M. P. Kierstan (1973). "Propapain and its conversion to papain: a new type of zymogen activation mechanism involving intramolecular thiol-disulphide interchange." Nat New Biol **242**(119): 167-70.

Brocklehurst, K. and J. P. Malthouse (1978). "Mechanism of the reaction of papain with substrate-derived diazomethyl ketones. Implications for the difference in site specificity of halomethyl ketones for serine proteinases and cysteine proteinases and for stereoelectronic requirements in the papain catalytic mechanism." Biochem J **175**(2): 761-4.

Brocklehurst, K., S. J. Willenbrock, *et al.* (1983). "Effects of conformational selectivity and of overlapping kinetically influential ionizations on the characteristics of pH-dependent enzyme kinetics. Implications of free-enzyme pKa variability in reactions of papain for its catalytic mechanism." Biochem J **211**(3): 701-8.

Brocklehurst, K., F. Willenbrock, *et al.* (1987). "Cysteine proteinases." Hydrolytic enzymes. A. Neuberger and K. Brocklehurst. Amsterdam, Elsevier Biomedical Press: 39-158.

Budt, K.H., A. Peyman, *et al.* (1995). "HIV protease inhibitor Hoe/By-793, structure-activity relationships in a series of C-2 symmetrical diols." Bioorg. Med. Chem. **3**: 559-571.

Buianouckas, F. R. (1995). "HIV results in the frame. HIV an illusion." Nature **375**(6528): 197; author reply 198.

Bukrinsky, M., K. Manogue, *et al.* (1995). "HIV results in the frame. Other approaches." Nature **375**(6528): 195-6; author reply 198.

Bullough, P. A., F. M. Hughson, *et al.* (1994). "Structure of influenza haemagglutinin at the pH of membrane fusion." Nature **371**(6492): 37-43.

Bulmer, M. (1988). "Codon usage and intragenic position." J Theor Biol **133**(1): 67-71.

Buseyne, F. and Y. Riviere (1993). "HIV-specific CD8+ T-cell immune responses and viral replication." Aids **7 Suppl 2**: S81-5.

Caflisch, A., H. J. Schramm, *et al.* (2000). "Design of dimerization inhibitors of HIV-1 aspartic proteinase: a computer-based combinatorial approach." J Comput Aided Mol Des **14**(2): 161-79.

Cai, M., Y. Huang, *et al.* (1998). "Solution structure of the His12 --> Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium." Protein Sci **7**(12): 2669-74.

Cai, Y. D., H. Yu, *et al.* (1998). "Artificial neural network method for predicting HIV protease cleavage sites in protein." J Protein Chem **17**(7): 607-15.

Cai, M., R. Zheng, *et al.* (1997). "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase [published erratum appears in Nat Struct Biol 1997 Oct;4(10):839-40]." Nat Struct Biol **4**(7): 567-77.

Cameron, C. E., M. Ghosh, *et al.* (1997). "Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand transfer efficiency are suppressed by HIV nucleocapsid protein." Proc Natl Acad Sci U S A **94**(13): 6700-5.

Capy, P., R. Vitalis, *et al.* (1996). "Relationships between transposable elements based upon the integrase- transposase domains: is there a common ancestor?" J Mol Evol **42**(3): 359-68.

Carr, C. M. and P. S. Kim (1993). "A spring-loaded mechanism for the conformational change of influenza hemagglutinin." Cell **73**(4): 823-32.

Carteau, S., S. C. Batson, *et al.* (1997). "Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+-dependent DNA integration *in vitro*." J Virol **71**(8): 6225-9.

Cassidy CS, Lin J, Frey PA. (1997). "A new concept for the mechanism of action of chymotrypsin: the role of the low-barrier hydrogen bond." Biochemistry. **36**(15):4576-84.

Chakrabarti, L., M. Guyader, *et al.* (1987). "Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses." Nature **328**(6130): 543-7.

Chan, A. Y., M. M. Chan, *et al.* (2002). "A dual protein expression system in Bacillus subtilis." Protein Expr Purif **26**(3): 337-42.

Chan, D. C., D. Fass, *et al.* (1997). "Core structure of gp41 from the HIV envelope glycoprotein." Cell **89**(2): 263-73.

Checroune, F., X. J. Yao, *et al.* (1995). "Incorporation of Vpr into human immunodeficiency virus type 1: role of conserved regions within the P6 domain of Pr55gag." J Acquir Immune Defic Syndr Hum Retrovirol **10**(1):1-7.

Chen, G. T. and M. Inouye (1994). "Role of the AGA / AGG codons, the rarest codons in global gene expression in *Escherichia coli*." Genes Dev **8**(21): 2641-52.

Chen J, and Y. Sun, (2003). "Modeling of the salt effects on hydrophobic adsorption equilibrium of protein." J Chromatogr A. **992**(1-2):29-40.

Chen, Z., Y. Li, *et al.* (1994). "Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases." J Biol Chem **269**(42): 26344-8.

Chen, Z., Y. Li, *et al.* (1995). "Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials." J Biol Chem **270**(37): 21433-6.

Cheng, Y. S., M. H. McGowan, *et al.* (1990). "High-level synthesis of recombinant HIV-1 protease and the recovery of active enzyme from inclusion bodies." Gene **87**(2): 243-8.

Cheng, Y. S., F. H. Yin, *et al.* (1990). "Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single-chain tethered dimer." Proc Natl Acad Sci U S A **87**(24): 9660-4.

Choe, H., M. Farzan, *et al.* (1996). "The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates." Cell **85**(7): 1135-48.

Chokekijchai, S., T. Shirasaka, *et al.* (1995). "In vitro anti-HIV-1 activity of HIV protease inhibitor KNI-272 in resting and activated cells: implications for its combined use with AZT or ddI." Antiviral Res **28**(1): 25-38.

Chou, J. J. (1993). "Predicting cleavability of peptide sequences by HIV protease via correlation-angle approach." J Protein Chem **12**(3): 291-302.

Chou, K. C. (1993). "A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins." J Biol Chem **268**(23): 16938-48.

Chou, K. C. and C. T. Zhang (1993). "Studies on the specificity of HIV protease: an application of Markov chain theory." J Protein Chem **12**(6): 709-24.

Chou, K. C., C. T. Zhang, *et al.* (1993). "A vector projection approach to predicting HIV protease cleavage sites in proteins." Proteins **16**(2): 195-204.

Chun, T. W., L. Carruth, *et al.* (1997). "Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection." Nature **387**(6629): 183-8.

Chun, T. W., R. T. Davey, Jr., *et al.* (1999). "Re-emergence of HIV after stopping therapy." Nature **401**(6756): 874-5.

Chun, T. W., D. Engel, *et al.* (1999). "Effect of interleukin-2 on the pool of latently infected, resting CD4+ T cells in HIV-1-infected patients receiving highly active anti-retroviral therapy." Nat Med **5**(6): 651-5.

Chun, T. W. and A. S. Fauci (1999). "Latent reservoirs of HIV: obstacles to the eradication of virus." Proc Natl Acad Sci U S A **96**(20): 10958-61.

Clapham, P. R. and R. A. Weiss (1997). "Immunodeficiency viruses. Spoilt for choice of co-receptors [news; comment]." Nature **388**(6639): 230-1.

Clarke, L. and J. Carbon (1976). "A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E.coli* genome." Cell **9**(1): 91-99.

Cleland, W. W. (1992). "Low-barrier hydrogen bonds and low fractionation factor bases in enzymatic reactions." Biochemistry **31**: 317-319.

Cleland W.W. and M.M. Kreevoy (1994). "Low-barrier hydrogen bonds and enzymic catalysis." Science. **264**(5167):1887-90.

Clever, J. L. and T. G. Parslow (1997). "Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation." J Virol **71**(5): 3407-14.

Coadou G,. J. Gharbi-Benarous, *et al.* (2003). "NMR studies of the phosphorylation motif of the HIV-1 protein Vpu bound to the F-box protein beta-TrCP." Biochemistry. **42**(50):14741-51.

Coates L., P.T. Erskine, *et al.* (2001). "A neutron Laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism." Biochemistry. **40**(44):13149-57.

Coates L., P.T. Erskine, *et al.*, (2003). "The structure of endothiapepsin complexed with the gem-diol inhibitor PD-135,040 at 1.37 A." Acta Crystallogr. D. Biol. Crystallogr. **59**(Pt 6):978-81. Epub 2003 May 23.

Coffin, J. M. (1992). "Genetic diversity and evolution of retroviruses." Curr Top Microbiol Immunol **176**: 143-64.

Coffin, J. M. (1995). "HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy." Science **267**(5197): 483-9.

Cohen, E. A., P. Gaudreau, *et al.* (1986) (1). "Neutralization of herpes simplex virus ribonucleotide reductase activity by an oligopeptide-induced antiserum directed against subunit H2." J Virol **60**(3): 1130-3.

Cohen, E. A., P. Gaudreau, *et al.* (1986) (2). "Specific inhibition of herpesvirus ribonucleotide reductase by a nonapeptide derived from the carboxy terminus of subunit 2." Nature **321**(6068): 441-3.

Cohen, E. A., R. A. Subbramanian, *et al.* (1996). "Role of auxiliary proteins in retroviral morphogenesis." Curr Top Microbiol Immunol **214**: 219-35.

Cohen, E.A., E.F. Terwilliger, *et al.* (1988). "Identification of a protein encoded by the vpu geneof HIV-1." Nature **334**:532-34.

Cohen, E.A., E.F. Terwilliger, *et al.* (1990). "Identification of HIV-1 vpr product and function." J. Acquir. Immune Defic. Syndr. **3**(1):11-8.

Cohen, J. (1995). "High turnover of HIV in blood revealed by new studies." Science **267**(5195): 179.

Cohen, O. J., G. Pantaleo, *et al.* (1996). "Antiretroviral monotherapy in early stage human immunodeficiency virus disease has no detectable effect on virus load in peripheral blood and lymph nodes." J Infect Dis **173**(4): 849-56.

Condra, J. H., W. A. Schleif, *et al.* (1995). "*In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors [see comments]." Nature **374**(6522): 569-71.

Conrad, B., R. S. Savchenko, *et al.* (1996). "A T7 promoter specific, inducible protein expression system for Bacillus subtilis." Mol Gen Genet **250**(2): 230-6.

Cordonnier, A., L. Montagnier, *et al.* (1989). "Single amino-acid changes in HIV envelope affect viral tropism and receptor binding." Nature **340**(6234): 571-4.

Cooper, J., S. Foundling, *et al.* (1987). "The structure of a synthetic pepsin inhibitor complexed with endothiapepsin." Eur J Biochem **169**(1):215-21.

Cooper, J.B. and D.A.A. Myles (2000). "A preliminary neutron Laue diffraction study of the aspartic proteinase endothiapepsin". Acta Cryst. **D56**: 246-248.

Cordonnier, A., Montagnier, L. and Emerman, M (1989). "Single amino-acid changes in HIV envelope affect viral tropism and receptor binding".Nature **340:** 571–574.

Cornish-Bowden, A. J., P. Greenwell, *et al.* (1969). "The rate determining step in pepsin-catalysed reactions, and evidence against an acyl-enzyme intermediate." Biochem. J. **113**(2): 369-75.

Cote, H. C., Z. L. Brumme, *et al.* (2001). "Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir." J Virol **75**(2): 589-94.

Craig, S. P., L. Yuan, *et al.* (1991). "High level expression in *Escherichia coli* of soluble, enzymatically active schistosomal hypoxanthine / guanine phosphoribosyltransferase and trypanosomal ornithine decarboxylase." Proc. Natl. Acad. Sci. U.SA. **88**: 2500-2504.

Crawford, S. and S. P. Goff (1985). "A deletion mutation in the 5' part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins." J Virol **53**(3): 899-907.

Creighton, T.E., (1989). Protein Structure, a practical approach. IRL Press, Oxford.

Crick, F., L. Barnett, *et al.* (1961). "General nature of the genetic code for proteins" Nature **192**: 1227

Cujec, T. P., H. Cho, *et al.* (1997). "The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme." Mol Cell Biol **17**(4): 1817-23.

D'Ettorre, C. and R. L. Levine (1994). "Reactivity of cysteine-67 of the human immunodeficiency virus-1 protease: studies on a peptide spanning residues 59 to 75." Arch Biochem Biophys **313**(1): 71-6.

Dalgleish, A. G., P. C. Beverley, *et al.* (1984). "The $CD4^+$ (T4) antigen is an essential component of the receptor for the AIDS retrovirus." Nature **312**(5996): 763-7.

Danley, D. E., K. F. Geoghegan, *et al.* (1989). "Crystallizable HIV-1 protease derived from expression of the viral pol gene in *Escherichia coli.*" Biochem Biophys Res Commun **165**(3): 1043-50.

Darke, P. L., C. T. Leu, *et al.* (1989). "Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease." J Biol Chem **264**(4): 2307-12.

Darke, P. L., R. F. Nutt, *et al.* (1988). "HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins." Biochem Biophys Res Commun **156**(1): 297-303.

Davies, D. R. (1990). "The structure and function of aspartic proteinases." Annu. Rev. Biophys. Biophy. Chem. **19**: 189-215.

Davis, D. A., A. A. Branca, *et al.* (1995). "Inhibition of the human immunodeficiency virus-1 protease and human immunodeficiency virus-1 replication by bathocuproine disulfonic acid Cu1+." Arch Biochem Biophys **322**(1): 127-34.

Davis, D. A., K. Dorsey, *et al.* (1996). "Regulation of HIV-1 protease activity through cysteine modification." Biochemistry **35**(7): 2482-8.

Davis, D. A., K. Yusa, *et al.* (1999). "Conserved cysteines of the human immunodeficiency virus type 1 protease are involved in regulation of polyprotein processing and viral maturation of immature virions." J Virol **73**(2): 1156-64.

Debouck, C., J. G. Gorniak, *et al.* (1987). "Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor." Proc Natl Acad Sci U S A **84**(24): 8903-6.

Dedera, D., W. Hu, *et al.* (1989)."Viral protein R of Human Immunodeficiency Virus types 1 and 2 is dispensable for replication and cytopathogenicity in lymphoid cells." J. Virol. Jul;**63**(7):3205-8.

De Mello, W.C. (2004). "Heart failure: how important is cellular sequestration? The role of the renin-angiotensin-aldosterone system." J. Mol. and Cell. Card. **37**;431-38

De Ronde, A., E. R. de Rooij, *et al.* (1996). "Zidovudine-resistant HIV strains in intravenous drug users and homosexual men in Amsterdam." Ned Tijdschr Geneeskd **140**(17): 932-4.

De Voss, J. J., Z. Sui, *et al.* (1994). "Haloperidol-based irreversible inhibitors of the HIV-1 and HIV-2 proteases." J Med Chem **37**(5): 665-73.

Dickson, C., Eisenman, R., Fan, H., Hunter, E. and Teich, N. (1984). "RNA Tumor Viruses." Molecular Biology of Tumor Viruses. Weiss, R., Teich, N., Varmus, H. and Coffin, J. New York, Cold Spring Harbor: 513 - 648.

DiIanni, C. L., L. J. Davis, *et al.* (1990). "Characterization of an active single polypeptide form of the human immunodeficiency virus type 1 protease." J Biol Chem **265**(28): 17348-54.

Divita, G., T. Restle, *et al.* (1994). "Inhibition of human immunodeficiency virus type 1 reverse transcriptase dimerization using synthetic peptides derived from the connection domain." J Biol Chem **269**(18): 13080-3.

Domingo, E., C. Escarmis, *et al.* (1996). "Basic concepts in RNA virus evolution." Faseb J **10**(8): 859-64.

Domingo, E. and J. J. Holland (1997). "RNA virus mutations and fitness for survival." Annu Rev Microbiol **51**: 151-78.

Dong, H., L. Nilsson and C. Kurland. (1996). "Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates" J. Mol. Biol. **260**(5): 649-63

Dorsey, B. D., R. B. Levin, *et al.* (1994). "L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor." J Med Chem **37**(21): 3443-51.

Drenth, J., J. N. Jansonius, *et al.* (1968). "Structure of Papain." Nature **218**: 929-932.

Drenth, J., J. N. Jansonius, *et al.* (1971). "The enzymes." P. D. Boyer. New York, Academic Press. **3:** 485. Drenth, J., K. H. Kalk, *et al.* (1976). Biochemistry **15**: 3731.

Dreyer, G. B., J. C. Boehm, *et al.* (1993). "A symmetric inhibitor binds HIV-1 protease asymmetrically." Biochemistry **32**(3): 937-47.

Dreyer, G.B., D.M. Lambert, *et al.* (1992). "Hydroxyethylene isosteres inhibitors of Human Immunodeficiency Virus-1 protease – structure activity analysis using enzyme-kinetics, X-ray crystallography, and infected T-Cell assays." Biochemistry **31**: 6646-6659.

Dreyer, G.B., B.W. Metcalf, *et al.* (1989). "Inhibition of Human Immunodeficiency Virus-1 protease *in vitro* – Rational Design of substrate-analog inhibitors." Proc. Natl. Acad. Sci. USA **86**: 9752-9756.

Dubay, J. W., S. J. Roberts, *et al.* (1992). "Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity." J Virol **66**(8): 4748-56.

Duesberg, P. and H. Bialy (1995). "HIV results in the frame. HIV an illusion." Nature **375**(6528): 197; author reply 198.

Dunn, B. M. and A. L. Fink (1984). "Cryoenzymology of porcine pepsin." Biochemistry **23**(22): 5241-7.

Dunn, B. M., A. Gustchina, *et al.* (1994). "Subsite preferences of retroviral proteinases." Methods Enzymol **241**: 254-78.

Dutia, B. M., M. C. Frame, *et al.* (1986). "Specific inhibition of herpesvirus ribonucleotide reductase by synthetic peptides." Nature **321**(6068): 439-41.

Dyda, F., A. B. Hickman, *et al.* (1994). "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases [see comments]." Science **266**(5193): 1981-6.

Earl, P. L., B. Moss, *et al.* (1991). "Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein." J Virol **65**(4): 2047-55.

Eastman, P. S., J. Mittler, *et al.* (1998). "Genotypic changes in human immunodeficiency virus type 1 associated with loss of suppression of plasma viral RNA levels in subjects treated with ritonavir (Norvir) monotherapy." J Virol **72**(6): 5154-64.

Ebbets-Reed, D., S. Scarlata, *et al.* (1996). "The major homology region of the HIV-1 gag precursor influences membrane affinity." Biochemistry **35**(45): 14268-75.

Edelhoch H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry. **6**(7):1948-54.

Ehrenberg, M. and C.G. Kurland. (1984). "Costs of accuracy determined by a maximal growth rate constraint" Q. Rev. Biophys. **17**(1): 45-82.

Embretson, J., M. Zupancic, *et al.* (1993). "Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS." Nature **362**(6418): 359-62.

Emerman, M. and Malim, H. (1998). "HIV-1 regulatory / accessory genes: keys to unraveling viral and host cell biology". Science **280**:1880-84.

Erickson, J. W. and S. K. Burt (1996). "Structural mechanisms of HIV drug resistance." Annu Rev Pharmacol Toxicol **36**: 545-71.

Erickson, J. W., S. V. Gulnik, *et al.* (1999). "Protease inhibitors: resistance, cross-resistance, fitness and the choice of initial and salvage therapies." Aids **13 Suppl A**: S189-204.

Erickson, J.W. and D. Kempf (1994). "Structure-based design of symmetric inhibitors of HIV-1 protease." Arch Virol Suppl **9**: 19-29.

Erickson, J.W., D. J. Neidhart, *et al.* (1990). "Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease." Science **249**(4968): 527-33.

Esnouf, R., J. Ren, *et al.* (1995). "Mechanism of inhibition of HIV-1 reverse transcriptase by non- nucleoside inhibitors." Nat Struct Biol **2**(4): 303-8.

Essex, M. and P. J. Kanki (1988). "The origins of the AIDS virus." Sci Am **259**(4): 64-71.

Ettmayer, P., M. Hübner, *et al.* (1994). "Novel extended transition state mimic in HIV-1 protease inhibitors with peripheral C2-symmetry." Bioorg. Med. Chem. Lett. **4**: 2851-2856.

Evans, D. T., D. H. O'Connor, *et al.* (1999). "Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef." Nat Med **5**(11): 1270-6.

Ewart, G.D., T. Sutherland, *et al.* (1996). "The Vpu protein of Human Immunodeficiency Virus type 1 forms cation-selective ion channels." J.Virol. **70**:7108-15.

Farmerie, W. G., D. D. Loeb, *et al.* (1987). "Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*." Science **236**(4799): 305-8.

Fass, D., R. A. Davey, *et al.* (1997). "Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution." Science **277**(5332): 1662-6.

Fauci, A. S. (1988). "The human immunodeficiency virus: infectivity and mechanisms of pathogenesis." Science **239**(4840): 617-22.

Fauci, A.S., Pantaleo G., *et al.* (1996). "HIV-1 protease inhibitors containing statine: inhibitory potency and antiviral activity." Ann. Intern. Med. **124** (7): 654-663.

Fehrentz, J. A., B. Chomier, *et al.* (1992). Biochem Biophys Res Commun **188**(2): 865-72.

Fehrentz, J.A., B. Chomier, *et al.* (1992). "Statine based tripeptides as potent inhibitors of HIV-1 replication." Biophys. Biochem. Res. Commun. **188** (2):873-878.

Feng, F., Broder, C. C., Kennedy, P. E. and Berger, E. A. (1996). "HIV-1-entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272,** 872–877.

Finkle, B. J. and E. L. Smith (1958). J. Biol. Chem. **230**: 669.

Finnegan, D. J. (1981). "Transposable elements and proviruses." Nature **292**(5826): 800-1.

Finnegan, D. J. (1983). "Retroviruses and transposable elements - which came first?" Nature **302**(5904): 105-6.

Finzi, D., M. Hermankova, *et al.* (1997). "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy." Science **278**(5341): 1295-300.

Fitzgerald, P. M., B. M. McKeever, *et al.* (1990). "Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution." J Biol Chem **265**(24): 14209-19.

Fitzgerald, P. M., B. M. McKeever, *et al.* (1990). "Crystallographic analysis of a complex between human immnodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution." J. Biol. Chem. **265**(24): 14209-19.

Flavell, A. (1981). "Did retroviruses evolve from transposable elements?" Nature **289**(5793): 10-1.

Forster, M.J., B. Mulloy, *et al.* (2000). "Molecular modelling study of HIV p17gag (MA) protein shell utilising data from electron microscopy and X-ray crystallography." J. Mol. Biol. **298**: 841-57.

Franchini, G., C. Gurgo, *et al.* (1987). "Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses." Nature **328**(6130): 539-43.

Franciskovitch, J., K. Houseman, *et al.* (1993). "A systematic evaluation of the inhibition of HIV-1 protease by its C- and N- terminal peptides." BioMed. Chem. Lett. **3**: 765-768.

Freed, E. O., G. Englund, *et al.* (1995). "Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection." J Virol **69**(6): 3949-54.

Frey P.A., S.A. Whitt, J.B. Tobin. (1994). "A low-barrier hydrogen bond in the catalytic triad of serine proteases." Science. **264**(5167):1927-30.

Fruton, J. S. (1976). "The mechanism of the catalytic action of pepsin and related acid proteinases." Adv. Enzymol. Relat. Areas Mol. Biol. **44**: 1-36.

Fujita, K., S. Omura, *et al.* (1997). "Rapid degradation of $CD4^+$ in cells expressing human immunodeficiency virus type 1 Env and Vpu is blocked by proteasome inhibitors [published erratum appears in J Gen Virol 1997 Aug;78(Pt 8):2129-30]." J Gen Virol **78**(Pt 3): 619-25.

Fukasawa, M., T. Miura, *et al.* (1988). "Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV / SIV group." Nature **333**(6172): 457-461.

Furfine, E. S., E. D'Souza, *et al.* (1992). "Two-step binding mechanism for HIV protease inhibitors." Biochemistry **31**(34): 7886-91.

Gallo, R. C. and L. Montagnier (1988). "AIDS in 1988 [see comments]." Sci Am **259**(4): 41-8.

Gallo, R. C., S. Z. Salahuddin, *et al.* (1984). "Frequent detection and isolation of cytopathic retrovirus (HTLVIII) from patients with AIDS and at risk for AIDS." Science **224**: 500-503.

Gallo, R., F. Wong-Staal, *et al.* (1988). "HIV/HTLV gene nomenclature [letter]." Nature **333**(6173): 504.

Gamble, T.R., K.R. Clauser, A.A. Kossiakoff (1994). "The production and X-ray structure determination of perdeuterated Staphylococcal nuclease." Biophys-Chem. **53**(1-2): 15-25.

Gamble, T. R., F. F. Vajdos, *et al.* (1996). "Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid." Cell **87**(7): 1285-94.

Gamble, T. R., S. Yoo, *et al.* (1997). "Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein." Science **278**(5339): 849-53.

Garcia-Martinez, L. F., G. Mavankal, *et al.* (1997). "Purification of a Tat-associated kinase reveals a TFIIH complex that modulates HIV-1 transcription." Embo J **16**(10): 2836-50.

Garman E.F., (1996). "Modern methods for rapid x-ray diffraction data collection from crystals of macromolecules." Methods Mol Biol. **56**:87-126.

Geiger, R., (1984). "Chemistry of the inhibitors of the renin - angiotensin system." Arzneimittelforschung. **34**(10B):1386-91.

Gerlt J.A., and P.G. Gassman (1993). "Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl transfer reactions, and displacement reactions of phosphodiesters." Biochemistry. **32**(45):11943-52.

Giam, C. Z. and I. Boros (1988). "*In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *Escherichia coli*." J Biol Chem **263**(29): 14617-20.

Gill SC, and P.H. von Hippel, (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Anal Biochem. **182**(2):319-26.

Gitti, R. K., B. M. Lee, *et al.* (1996). "Structure of the amino-terminal core domain of the HIV-1 capsid protein." Science **273**(5272): 231-5.

Glazer, A. N. and E. L. Smith (1971). "The enzymes." P. D. Boyer. New York, Academic Press. **3:** 501.

Goedert, J. J. and R. C. Gallo (1985). "Epidemiological evidence that HTLV-III is the AIDS agent." Eur J Epidemiol **1**(3): 155-9.

Goldman, E., A. H. Rosenberg, *et al.* (1995). "Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*." J. Mol. Biol. **245**(5): 467-473.

Gonda, M. A., F. Wong-Staal, *et al.* (1985). "Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus." Science **227**(4683): 173-7.

Goudsmit, J., A. de Ronde, *et al.* (1997). "Broad spectrum of in vivo fitness of human immunodeficiency virus type 1 subpopulations differing at reverse transcriptase codons 41 and 215." J Virol **71**(6): 4479-84.

Graham, R. W., T. Atkinson, *et al.* (1993). "Rational design and PCR-based synthesis of an artificial Schizophyllum commune xylanase gene." Nucleic Acids Res **21**(21): 4923-8.

Grant, S. K., I. C. Deckman, *et al.* (1992). "Use of protein unfolding studies to determine the conformational and dimeric stabilities of HIV-1 and SIV proteases." Biochemistry **31**(39): 9491-501.

Graves, B. J., M. H. Hatada, *et al.* (1991). "The three-dimensional x-ray crystal structure of HIV-1 protease complexed with a hydroxyethylene inhibitor." Adv Exp Med Biol **306**: 455-60.

Graves, M. C., J. J. Lim, *et al.* (1988). "An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity." Proc Natl Acad Sci U S A **85**(8): 2449-53.

Griffiths, J. T., L. H. Phylip, *et al.* (1992). "Different requirements for productive interaction between the active site of HIV-1 proteinase and substrates containing -hydrophobic*hydrophobic- or -aromatic*pro- cleavage sites." Biochemistry **31**(22): 5193-200.

Guatelli, J. C. (1997). "The positive influence of Nef on viral infectivity." Res Virol **148**(1): 34-7.

Gulnik, S. V., L. I. Suvorov, *et al.* (1995). "Kinetic characterisation and cross resistance patterns of HIV-1 protease mutants selected under drug pressure." Biochemistry **34**(29): 9282-7.

Guo, J., L. E. Henderson, *et al.* (1997). "Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA." J Virol **71**(7): 5178-88.

Gustafson, M. E., K. D. Junger, *et al.* (1995). "Large scale production of HIV-1 proteasefrom *Escherichia coli* using selective extractionand membrane fractionation." Protein Expression and Purification **6**: 512-518.

Gustchina, A., J. Kervinen, *et al.* (1996). "Structure of equine infectious anemia virus proteinase complexed with an inhibitor." Protein Sci **5**(8): 1453-65.

Guyader, M., M. Emerman, *et al.* (1987). "Genome organization and transactivation of the human immunodeficiency virus type 2." Nature **326**(6114): 662-9.

Haase, A. T. (1994). "The role of active and covert infections in lentivirus pathogenesis." Ann N Y Acad Sci **724**: 75-86.

Haase, A. T. (1999). "Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues." Annu Rev Immunol **17**: 625-56.

Haase, A. T., K. Henry, *et al.* (1996). "Quantitative image analysis of HIV-1 infection in lymphoid tissue." Science **274**(5289): 985-9.

Hansen, J., S. Billich, *et al.* (1988). "Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody." Embo J 7(6): 1785-91.

Hanson JC, Schoenborn BP. (1981). "Real space refinement of neutron diffraction data from sperm whale carbonmonoxymyoglobin." J. Mol Biol. **153**(1):117-46.

Haren, L., B. Ton-Hoang, *et al.* (1999). "Integrating DNA: transposases and retroviral integrases." Annu Rev Microbiol **53**: 245-81.

Harrison, R. W. and I. T. Weber (1994). "Molecular dynamics simulations of HIV-1 protease with peptide substrate." Protein Eng **7**(11): 1353-63.

Harte, W. E. and D. L. Beveridge (1994). "Probing structure-function relationships in Human Immunodeficiency Virus Type 1 protease via molecular dynamics simulation." Retroviral Proteases. L. C. Kuo and J. A. Shafer. London, Academic Press Ltd. **241:** 178-194.

Harte, W. E., Jr., S. Swaminathan, *et al.* (1990). "Domain communication in the dynamical structure of human immunodeficiency virus 1 protease." Proc Natl Acad Sci U S A **87**(22): 8864-8.

Hartung, S., K. Boller, *et al.* (1992). "Quantitation of a lentivirus i its natural host: simian immunodeficiency virus in African green monkeys." J. Virol. **66**(4): 2143-9.

Haseltine, W. A. (1991). "Molecular biology of the human immunodeficiency virus type 1." Faseb J **5**(10): 2349-60.

Hehlmann, R., R. Brack-Werner, *et al.* (1988). "Human endogenous retroviruses." Leukemia **2**(12 Suppl): 167S-177S.

Heimbach, J.C., V.M. Garsky, *et al.* (1989). "Affinity purification of the HIV-1 protease. " Biochem Biophys Res Commun **164**(3): 955-60.

Hellen, C. U., H. G. Krausslich, *et al.* (1989). "Proteolytic processing of polyproteins in the replication of RNA viruses." Biochemistry **28**(26): 9881-90.

Henderson, L. E., H. C. Krutzsch, *et al.* (1983). "Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification." Proc Natl Acad Sci U S A **80**(2): 339-43.

Hill, C. P., D. Worthylake, *et al.* (1996). "Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly." Proc Natl Acad Sci U S A **93**(7): 3099-104.

Hinks, J. A., M. Roe, *et al.* (2003). "Expression, purification, and preliminary X-ray analysis of the tandem BRCT domains from Rhp9 / Crb2." Acta. Cryst. **D59**: 1230-1233.

Hirsch, V.M., P.R. Johnson. (1994). "Pathogenic diversity of simian immunodeficiency viruses." Virus Res. **32**:183-203.

Ho, D. D. and Y. Huang (2002). "The HIV-1 vaccine race." Cell **110**(2): 135-8.

Ho, D. D., A. U. Neumann, *et al.* (1995). "Rapid turnover of plasma virions and $CD4^+$ lymphocytes in HIV-1 infection [see comments]." Nature **373**(6510): 123-6.

Hofmann, T., A. L. Fink, *et al.* (1984). "Cryoenzymology of penicillopepsin; with an appendix; mechanism of action of aspartyl proteinases." Biochemistry **23**(22): 5247-56.

Hofmann, T. and R. S. Hodges (1982). "A new chromophoric substrate for penicillopepsin and other fungal aspartic proteinases." Biochem. J. **203**(3): 603-610.

Holland, J. J. and E. D. Kiehn (1968). "Specific cleavage of viral proteins as steps in the synthesis and maturation of enteroviruses." Proc Natl Acad Sci U S A **60**(3): 1015-22.

Hong, L., A. Treharne, *et al.* (1996). "Crystal structures of complexes of a peptidic inhibitor with wild-type and two mutant HIV-1 proteases." Biochemistry **35**(33): 10627-33.

Hope, T. J. (1997). "Viral RNA export." Chem Biol **4**(5): 335-44.

Horton R, P. Spearman, *et al.* (1994). "HIV-2 viral protein X association with the GAG p27 capsid protein." Virology. Mar;**199**(2):453-7.

Hsiou, Y., J. Ding, *et al.* (1996). "Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms." Structure **4**(7): 853-60.

Hsu, I. N., L. T. Delbaere, *et al.* (1977). "Penicillopepsin: 2.8 A structure, active site conformation and mechanistic implications." Adv. Exp. Med. Biol. **95**: 61-81.

Huang, J., R. C. Newton, *et al.* (1987). "High level expression in *Escherichia coli* of a soluble and fully active recombinant interleukin-1 beta." Mol. Biol. Med. **4**(3): 169-181.

Huang, M., J. M. Orenstein, *et al.* (1995). "p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease." J Virol **69**(11): 6810-8.

Huang, X., J. J. Barchi, Jr., *et al.* (1997). "Glycosylation affects both the three-dimensional structure and antibody binding properties of the HIV-1IIIB GP120 peptide RP135." Biochemistry **36**(36): 10846-56.

Huang, Y., J. Wang, *et al.* (1997). "Primer tRNA3Lys on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1." J Virol **71**(1): 726-8.

Huet, T.R., A. Cheynier, *et al.* (1990). "Genetic organisation of a Chimpanzee lentivirus related to HIV-1." Nature **345**:356-59.

Hur O., C. Leja, and M.F. Dunn. (1996). "Evidence of a low-barrier hydrogen bond in the tryptophan synthase catalytic mechanism." Biochemistry. **35**(23):7378-86.

Ikemura, T. (1981). "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes." J. Mol. Biol. **146**(1): 1-21.

Jacobo-Molina, A., J. Ding, *et al.* (1993). "Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA." Proc Natl Acad Sci U S A **90**(13): 6320-4.

Jacobson, M. F. and D. Baltimore (1968). "Polypeptide cleavages in the formation of poliovirus proteins." Proc Natl Acad Sci U S A **61**(1): 77-84.

James, J. S. (1995). "New information on HIV rapid turnover--what does it mean?" AIDS Treat News **215**: 1-4.

James, M., A. Sielecki, *et al.* (1982). "Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin." Proc Natl Acad Sci U S A **79**(20): 6137-41.

James, M. N., I. N. Hsu, *et al.* (1977). "Mechanism of acid protease catalysis based on the crystal structure of penicillopepsin." Nature **267**(5614): 808-13.

James, M. N. and A. R. Sielecki (1983). "Structure and refinement of penicillopepsin at 1.8 A resolution." J Mol Biol **163**(2): 299-361.

James, M. N. and A. R. Sielecki (1986). "Molecular structure of an apsartic proteinase zymogen, porcine pepsinogen, at 1.8 A resolution." Nature **319**(6048): 33-8.

Jaskolski, M., A. G. Tomasselli, *et al.* (1991). "Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene- based inhibitor." Biochemistry **30**(6): 1600-9.

Jenkins, J., I. Tickle, *et al.* (1977). "X-ray analysis and circular dichroism of the acid protease from Endothia parasitica and chymosin." Adv. Exp. Med. Biol. **95**: 43-60.

Jiang, S. and A. K. Debnath (2000). "Development of HIV entry inhibitors targeted to the coiled-coil regions of gp41." Biochem Biophys Res Commun **269**(3): 641-6.

Jiang, S., K. Lin, *et al.* (1993). "HIV-1 inhibition by a peptide." Nature **365**(6442): 113.

Jones, D. H. and B. H. Howard (1990). "A rapid method for site-specific mutagenesis and directional subcloning by using the polymerase chain reaction to generate recombinant circles." Biotechniques **8**(2): 178-83.

Jones, K. A. (1997). "Taking a new TAK on tat transactivation [comment]." Genes Dev **11**(20): 2593-9.

Judice, J. K., J. Y. Tom, *et al.* (1997). "Inhibition of HIV type 1 infectivity by constrained alpha-helical peptides: implications for the viral fusion mechanism." Proc Natl Acad Sci U S A **94**(25): 13426-30.

Jungheim, L. N., T. A. Shepherd, *et al.* (1996). "Potent human immunodeficiency virus type 1 protease inhibitors that utilize noncoded D-amino acids as P2/P3 ligands." J Med Chem **39**(1): 96-108.

Kaldor, S. W., V. J. Kalish, *et al.* (1997). "Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease." J Med Chem **40**(24): 3979-85.

Karlstrom, A. R. and R. L. Levine (1991). "Copper inhibits the protease from human immunodeficiency virus 1 by both cysteine-dependent and cysteine-independent mechanisms." Proc Natl Acad Sci U S A **88**(13): 5552-6.

Karlstrom, A. R., B. D. Shames, *et al.* (1993). "Reactivity of cysteine residues in the protease from human immunodeficiency virus: identification of a surface-exposed region which affects enzyme function." Arch Biochem Biophys **304**(1): 163-9.

Kay, J., E. G. Afting, *et al.* (1982). "The effects of lactoyl-pepstatin and the pepsin inhibitor peptide on pig cathepsin D." Biochem J **203**(3): 795-7.

Kempf, D. J., K. C. Marsh, *et al.* (1995). "ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans." Proc Natl Acad Sci U S A **92**(7): 2484-8.

Kempf, D. J., R. A. Rode, *et al.* (1998). "The duration of viral suppression during protease inhibitor therapy for HIV-1 infection is predicted by plasma HIV-1 RNA at the nadir." Aids **12**(5): F9-14.

Kempf, D. J., H. L. Sham, *et al.* (1998). "Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy." J Med Chem **41**(4): 602-17.

Kiriyama, A., T. Nishiura, *et al.* (1996). "Binding characteristics of KNI-272 to plasma proteins, a new potent tripeptide HIV protease inhibitor." Biopharm Drug Dispos **17**(9): 739-51.

Kiso, Y. (1996). "Design and synthesis of substrate-based peptidomimetic human immunodeficiency virus protease inhibitors containing the hydroxy-methyl-carbonyl isostere." Biopolymers **40**(2): 235-44.

Kiso, Y., S. Yamaguchi, *et al.* (1998). "KNI-577, a potent small-sized HIV protease inhibitor based on the dipeptide containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic." Arch Pharm (Weinheim) **331**(3): 87-9.

Klabe, R. M., L. T. Bacheler, *et al.* (1998). "Resistance to HIV protease inhibitors: a comparison of enzyme inhibition and antiviral potency." Biochemistry **37**(24): 8735-42.

Kowalski, M., J. Potz, *et al.* (1987). "Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1." Science **237**(4820): 1351-5.

Krohn, A., S. Redshaw, *et al.* (1991). "Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere." J Med Chem **34**(11): 3340-2.

Kahyaoglu A., K. Haghjoo, *et al.* (1997). "Low barrier hydrogen bond is absent in the catalytic triads in the ground state but is present in a transition-state complex in the prolyl oligopeptidase family of serine proteases." J. Biol. Chem. **272**(41):25547-54.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*." Curr Opin Biotechnol **6**(5): 494-500.

Kanki, P. J., J. Alroy, *et al.* (1985). "Isolation of T-lymphotropic retrovirus related to HTLV-III / LAV from wild caught African green monkeys." Science **230**(4728): 951-954.

Kaplan, J. E., T. J. Spira, *et al.* (1996). "14-year follow-up of HIV-infected homosexual men with lymphadenopathy syndrome." J Acquir Immune Defic Syndr Hum Retrovirol **11**(2): 206-8.

Katoh, I., Y. Yoshinaka, *et al.* (1985). "Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity." Virology **145**(2): 280-92.

Kempf, D.J. (1994). "Design of symmetry-based, peptidomimetic inhibitors of Human Immunodeficiency Virus Protease". Methods Enzymol. **241**:234-254.

Kempf, D.J., K.C. Marsh, *et al.* (1995). "ABT-538 is a potent inhibitor of Human Immunodeficiency Virus protease and has high oral bioavailability in humans." Proc. Natl. Acad. Sci. USA **92**: 2484-2488.

Kempf, D.J., H.L. Sham (1996). "HIV Protease inhibitors." Curr. Pharm. Design **2**:225-246.

Kerkau, T., I. Bacik, *et al.* (1997). "The human immunodeficiency virus type 1 (HIV-1) Vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules." J Exp Med **185**(7): 1295-305.

Kimmel, J. R. and E. L. Smith (1954). "Crystalline papain. I. Preparation, specificity, and activation." J. Biol. Chem. **207**: 515-531.

Kimmel, J. R. and E. L. Smith (1957). "The properties of Papain." Advances in Enzymology **19**: 267-334.

Klein, I. B. and J. F. Kirsch (1969). "The mechanism of the activation of papain." Biochem Biophys Res Commun **34**(5): 575-81.

Kleinert, H.D., (1989). "Renin inhibitors: discovery and development. An overview and perspective." Am. J. Hypertens. **2**(10):800-8.

Klimkait, T., K. Strebel, *et al.* (1990). "The Human Immunodeficiency Virus type 1-specific protein Vpu is reuired for efficient virus maturation and release." J.Virol. **64**:621-29.

Kohl, N. E., E. A. Emini, *et al.* (1988). "Active human immunodeficiency virus protease is required for viral infectivity." Proc Natl Acad Sci U S A **85**(13): 4686-90.

Kohlstaedt, L. A., J. Wang, *et al.* (1992). "Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor." Science **256**(5065): 1783-90.

Kondo, E. and H. G. Gottlinger (1996). "A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr." J Virol **70**(1): 159-64.

Konvalinka, J., P. Strop, *et al.* (1990). "Sub-site preferences of the aspartic proteinase from the human immunodeficiency virus, HIV-1." FEBS Lett **268**(1): 35-8.

Korant, B. D. and C. J. Rizzo (1991). "An E. coli expression system which detoxifies the HIV protease." Biomed Biochim Acta **50**(4-6): 643-6.

Kossiakoff, A.A. and S.A. Spencer (1981). "Direct determination of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine protease: neutron structure of trypsin." Biochemistry **20**: 6462-6474.

Kowalski, M., J. Potz, *et al.* (1987). "Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1." Science **237**(4820): 1351-5.

Kramer, R. A., M. D. Schaber, *et al.* (1986). "HTLV-III gag protein is processed in yeast cells by the virus pol- protease." Science **231**(4745): 1580-4.

Kraus, G., A. Werner, *et al.*(1989). "Isolation of human immunodeficiency virus-related simian immunodeficiency viruses from African green monkeys." Proc. Natl. Acad. Sci. USA. **86**(8): 2892-6.

Krausslich, H. G., R. H. Ingraham, *et al.* (1989). "Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides." Proc Natl Acad Sci U S A **86**(3): 807-11.

Krohn, A., S. Redshaw, *et al.* (1991). "Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere." J Med Chem **34**(11): 3340-2.

Kurth, R., G. Kraus, *et al.* (1988). "AIDS: animal retrovirus models and vaccines." J. Acquired Immune Defic. Syndr. **1**: 284-294.

Kwong, P. D., R. Wyatt, *et al.* (1998). "Structure of an HIV gp120 envelope glycoprotein in complex with the $CD4^+$ receptor and a neutralizing human antibody [see comments]." Nature **393**(6686): 648-59.

Lam, P. Y., P. K. Jadhav, *et al.* (1994). "Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors." Science **263**(5145): 380-4.

Lam, P. Y., Y. Ru, *et al.* (1996). "Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas." J Med Chem **39**(18): 3514-25.

Larder, B. A., G. Darby, *et al.* (1989). "HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy." Science **243**(4899): 1731-4.

Larder, B. A. and S. D. Kemp (1989). "Multiple mutations in HIV-1 reverse transcriptase confer high level resistance to Zidovudine (AZT)." Science **246**: 1155-1158.

Laughrea, M., L. Jette, *et al.* (1997). "Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization." J Virol **71**(5): 3397-406.

Lebon, F., N. Boggetto, *et al.* (2002). "Metal-organic compounds: a new approach for drug discovery. N1-(4-methyl-2-pyridyl)-2,3,6-trimethoxybenzamide copper(II) complex as an inhibitor of human immunodeficiency virus 1 protease." Biochem Pharmacol **63**(10): 1863-73.

Lebon, F., E. deRosny, *et al.* (1998). "de novo drug design of a new copper chelate molecule acting as HIV-1 protease inhibitor." Eur. J. Med. Chem. **33**: 1-5.

Lebon, F. and M. Ledecq (2000). "Approaches to the design of effective HIV-1 protease inhibitors." Curr Med Chem **7**(4): 455-77.

Lebon, F., M. Ledecq, *et al.* (1999). "Synthesis and structural analysis of copper(II) pyridine amide complexes as HIV-1 protease inhibitors." Journal of the Chemical society. Perkin transactions. **2**(4): 795-800.

Lebon, F., M. Ledecq, *et al.* (2001). "Synthesis and structural analysis of the copper(II) complexes of N2-(2-pyridylmethyl)-2-pyridinecarboxamide." J Inorg Biochem **86**(2-3): 547-54.

Leckie, B. J., M. Szelke, *et al.* (1985). "Human renin inhibitors." Biochem Soc Trans **13**(6): 1029-32.

Lee, C. S., N. Choy, *et al.* (1996). "Design, synthesis, and characterization of dipeptide isostere containing cis-epoxide for the irreversible inactivation of HIV protease." Bioorg. Med. Chem. Lett. **6**: 589-594.

Le Gall, S., J. M. Heard, *et al.* (1997). "Analysis of Nef-induced MHC-I endocytosis." Res Virol **148**(1): 43-7.

Le Grice, S. F., J. Mills, *et al.* (1988). "Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation." Embo J **7**(8): 2547-53.

Leonard, C. K., M.W. Spellman, *et al.* (1990). "Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cell." J. Biol.Chem. **265** (18): 10373–10382.

Levy, J. A., B. Ramachandran, *et al.* (1996). "Plasma viral load, CD4+ cell counts, and HIV-1 production by cells." Science **271**(5249): 670-1.

Lifson, J. D., M. B. Feinberg, *et al.* (1986). "Induction of $CD4^+$-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein." Nature **323**(6090): 725-8.

Lin, C. and W.H. Frishman (1996). "Renin inhibition: a novel therapy for cardiovascular disease." Am. Heart J. **131**(5):1024-34.

Lin, X. L., Y. Z. Lin, *et al.* (1994). "Relationships of human immunodeficiency virus protease with eukaryotic aspartic proteases." Methods Enzymol **241**: 195-224.

Liuzzi, M., R. Deziel, *et al.* (1994). "A potent peptidomimetic inhibitor of HSV ribonucleotide reductase with antiviral activity in vivo." Nature **372**(6507): 695-8.

Lodi, P. J., J. A. Ernst, *et al.* (1995). "Solution structure of the DNA binding domain of HIV-1 integrase." Biochemistry **34**(31): 9826-33.

Loeb, D. D., C. A. d. Hutchison, *et al.* (1989). "Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases." J Virol **63**(1): 111-21.

Loshon, C. A., F. Tovar-Rojo, *et al.* (1989). "The expression of a highly expressed Bacillus subtilis gene is not reduced by introduction of multiple codons normally not present in such genes." FEMS Microbiol Lett **53**(1-2): 59-63.

Lowe, G. (1970). "The structure and mechanism of action of papain." Philos Trans R Soc Lond B Biol Sci **257**(813): 237-48.

Lu, M., S. C. Blacklow, *et al.* (1995). "A trimeric structural domain of the HIV-1 transmembrane glycoprotein." Nat Struct Biol **2**(12): 1075-82.

Lu, Y. L., R. P. Bennett, *et al.* (1995). "A leucine triplet repeat sequence (LXX)4 in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles." J Virol **69**(11): 6873-9.

Luther, R.R., H.H. Stein, *et al.* (1989). "Renin inhibitors:specific modulators of the renin-angiotensin system." Arzneimittelforschung. **39**(1):1-5.

Maizel, J. V., Jr. and D. F. Summers (1968). "Evidence for differences in size and composition of the poliovirus- specific polypeptides in infected HeLa cells." Virology **36**(1): 48-54.

Majors, J. E., R. Swanstrom, *et al.* (1981). "DNA intermediates in the replication of retroviruses are structurally (and perhaps functionally) related to transposable elements." Cold Spring Harb Symp Quant Biol **45**(Pt 2): 731-8.

Mammano, F., E. Kondo, *et al.* (1995). "Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope glycoproteins with short cytoplasmic domains." J Virol **69**(6): 3824-30.

Mammano, F., C. Petit, *et al.* (1998). "Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients." J Virol **72**(9): 7632-7.

Mangasarian, A. and D. Trono (1997). "The multifaceted role of HIV Nef." Res Virol **148**(1): 30-3.

Mansky, L. M. and H. M. Temin (1995). "Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase." J Virol **69**(8): 5087-94.

Marciniszyn, J. J., J. A. Hartsuck, *et al.* (1976). "Mode of inhibition of acid proteases by pepstatin." J. Biol. Chem. **251**(22): 7088-94.

Margottin, F., S.P. Bour, *et al.* (1998). "A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif." Mol Cell. **1**(4):565-74.

Markowitz, M., H. Mo, *et al.* (1995). "Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor." J Virol **69**(2): 701-6.

Marlink, R. G., D. Ricard, *et al.* (1988). "Clinical, hematologic, and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type-2 (HIV-2)." AIDS Res Hum Retroviruses **4**(2): 137-48.

Martin, S. L., B. Vrhovski, *et al.* (1995). "Total synthesis and expression in *Escherichia coli* of a gene encoding human tropoelastin." Gene **154**(2): 159-66.

Mason S., G.A. Bentley, and G. McIntyre (1984). "Deuterium exchange in lysozyme at 1.4 Å resolution" Neutrons in Biology (Schoenborn, B., Ed.) 323-334, Plenum, New York.

Massiah, M. A., D. Worthylake, *et al.* (1996). "Comparison of the NMR and X-ray structures of the HIV-1 matrix protein: evidence for conformational changes during viral assembly." Protein Sci **5**(12): 2391-8.

McCune, J. M., L. B. Rabin, *et al.* (1988). "Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus." Cell **53**(1): 55-67.

McKeever, B. M., M. A. Navia, *et al.* (1989). "Crystallization of the aspartylprotease from the human immunodeficiency virus, HIV-1." J Biol Chem **264**(4): 1919-21.

McPhee, F., A. C. Good, *et al.* (1996). "Engineering human immunodeficiency virus 1 protease heterodimers as macromolecular inhibitors of viral maturation." Proc Natl Acad Sci U S A **93**(21): 11477-81.

Meek, T. D., B. D. Dayton, *et al.* (1989). "Human immunodeficiency virus 1 protease expressed in *Escherichia coli* behaves as a dimeric aspartic protease." Proc Natl Acad Sci U S A **86**(6): 1841-5.

Meek, T.D. D.M. Lambert, *et al.* (1990). "Inhibitions of HIV-1 protease in infected lymphocytes-T by synthetic peptide analogues." Nature **343**: 90-92.

Meek, T. D., E. J. Rodriguez, *et al.* (1994). "Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral proteases. Retroviral Proteases." L. C. Kuo and J. A. Shafer. London, Academic Pres Ltd. **241:** 127-156.

Mellors, J. W., R. F. Schinazi, *et al.* (1996). "Mutations in Retroviral Genes Associated with Drug Resistance." Human Retroviruses and AIDS. Myers ed. Los Alamos, Los AlamosNational Laboratory. III: 206.

Melnick, M., S. H. Reich, *et al.* (1996). "Bis tertiary amide inhibitors of the HIV-1 protease generated via protein structure-based iterative design." J Med Chem **39**(14): 2795-811.

Menard, R., E. Carmona, *et al.* (1993). "The specificity of the S1' subsite of cysteine proteases." FEBS Letters **328**: 107-110.

Menendez-Arias, L., M. Young, *et al.* (1992). "Purification and characterization of the mouse mammary tumor virus protease expressed in *Escherichia coli*." J Biol Chem **267**(33): 24134-9.

Merluzzi, V. J., K. D. Hargrave, *et al.* (1990). "Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor." Science **250**(4986): 1411-3.

Mervis, R. J., N. Ahmad, *et al.* (1988). "The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame,

identification of posttranslational modifications, and evidence for alternative gag precursors." J Virol **62**(11): 3993-4002.

Michie, C. (1995). "HIV results in the frame. Toxic shock." Nature **375**(6528): 197-8; author reply 198.

Miller, M. D., Y. C. Bor, *et al.* (1995). "Target DNA capture by HIV-1 integration complexes." Curr Biol **5**(9): 1047-56.

Miller, M., M. Geller, *et al.* (1997). "Analysis of the structure of chemically synthesized HIV-1 protease complexed with a hexapeptide inhibitor. Part I: Crystallographic refinement of 2 A data." Proteins **27**(2): 184-94.

Miller, M., M. Jaskolski, *et al.* (1989). "Crystal structure of a retroviral protease proves relationship to aspartic protease family." Nature **337**(6207): 576-9.

Miller, M., J. Schneider, *et al.* (1989). "Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution." Science **246**(4934): 1149-52.

Mills, H. R., H. A. Overton, *et al.* (1990). "Expression of HIV-1 gag and pol gene products using recombinant baculoviruses." Retroviral proteases. L. H. Pearl. London, Macmillan Press Ltd.: 107-115.

Mimoto, T., J. Imai, *et al.* (1991). "Rational design and synthesis of a novel class of active site targeted HIV protease inhibitors containing a hydroxyethylcarbonyl isosteres – use of phenylnorstatine or allophenylnorststine as transition state mimic." Chem. Pharm. Bull. (Tokyo). **39**(9): 2465-7.

Mimoto, T., R. Kato, *et al.* (1999). "Structure-activity relationship of small-sized HIV protease inhibitors containing allophenylnorstatine." J Med Chem **42**(10): 1789-802.

Mitsuya, H., R. Yarchoan, *et al.* (1990). "Molecular targets for AIDS therapy." Science **249**(4976): 1533-44.

Mo, H., M. Markowitz, *et al.* (1996). "Design, synthesis, and resistance patterns of MP-134 and MP-167, two novel inhibitors of HIV type 1 protease." AIDS Res Hum Retroviruses **12**(1): 55-61.

Moebius, U., L. Clayton, *et al.* (1992). "The human immunodeficiency virus gp120 binding site of CD4: Delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high resolution CD4 atomic structure." J. Exp. Med. **176:** 507–517.

Moir, S., J. Perreault, and L. Poulin, (1996). "Postbinding events mediated by human immunodeficiency virus type 1 are sensitive to modifications in the D4 transmembrane linked region of CD4." J. Virol. **70:** 8019–8028.

Mous, J., E. P. Heimer, *et al.* (1988). "Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in *Escherichia coli*." J Virol **62**(4): 1433-6.

Molla, A., M. Korneyeva, *et al.* (1996). "Ordered accumulation of mutations in HIV protease confers resistance to ritonavir." Nat Med **2**(7): 760-6.

Momany, C., L. C. Kovari, *et al.* (1996). "Crystal structure of dimeric HIV-1 capsid protein." Nat Struct Biol **3**(9): 763-70.

Moore, J. P. (1997). "Coreceptors: implications for HIV pathogenesis and therapy". Science **276:** 51–52.

Moore, M.L., W.M. Bryan, *et al.* (1989). "Peptide-substrates and inhibitors of the HIV-1 Protease." Biochem. Biophys. Res. Commun. **159**: 420-425.

Moore, J. P., J.A. McKeating, *et al.* (1990). "Dissociation of gp120 from HIV-1 virions induced by soluble CD4". Science **250:** 1139–1142.

Morellet, N., N. Jullian, *et al.* (1992). "Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR." Embo J **11**(8): 3059-65.

Morihara, K. and T. Oka (1981). "Peptide based synthesis catalysed by subtilisin. papain and pepsin." J. Biochem. (Tokyo) **89**(2): 385-95.

Morishima, H., T. Takita, *et al.* (1970). "The structure of pepstatin." J Antibiot (Tokyo) **23**(5): 263-5.

Mosier, D. E. (1995). "HIV results in the frame. CD4+ cell turnover." Nature **375**(6528): 193-4; author reply 198.

Mueller, B. U., B. D. Anderson, *et al.* (1998). "Pharmacokinetics of the protease inhibitor KNI-272 in plasma and cerebrospinal fluid in nonhuman primates after intravenous dosing and in human immunodeficiency virus-infected children after intravenous and oral dosing." Antimicrob Agents Chemother **42**(7): 1815-8.

Mulichak, A. M., J. O. Hui, *et al.* (1993). "The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors." J Biol Chem **268**(18): 13103-9.

Murao, S. and S. Satoi (1970). "New pepsin inhibitor (S-PI) from Streptomyces EF-44-201." Agric. Biol. Chem. **34**: 1265-1267.

Nakamura, Y., K. Wada, *et al.* (1996). "Codon usage tabulated from the international DNA sequence databases." Nucleic Acids Res. **24**(1) :214-5.

Navia, M. A., P. M. Fitzgerald, *et al.* (1989). "Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1." Nature **337**(6208): 615-20.

Nicholson, L. K., T. Yamazaki, *et al.* (1995). "Flexibility and function in HIV-1 protease." Nat Struct Biol **2**(4): 274-80.

Nijhuis, M., R. Schuurman, *et al.* (1999). "Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy." Aids **13**(17): 2349-59.

Norley, S.G. (1996). "SIVagm infection of its natural African green monkey host." Immunol. Lett. **51**(1-2): 53-8.

Novella, I. S., D. K. Clarke, *et al.* (1995). "Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells." J Virol **69**(11): 6805-9.

Novella, I. S., E. A. Duarte, *et al.* (1995). "Exponential increases of RNA virus fitness during large population transmissions." Proc Natl Acad Sci U S A **92**(13): 5841-4.

Nowak, M. A. (1995). "AIDS pathogenesis: from models to viral dynamics in patients." J Acquir Immune Defic Syndr Hum Retrovirol **10**(Suppl 1): S1-5.

Nowak, M. A., R. M. Anderson, *et al.* (1991). "Antigenic diversity thresholds and the development of AIDS." Science **254**(5034): 963-9.

Nowak, M. A., S. Bonhoeffer, *et al.* (1995). "HIV results in the frame. Results confirmed." Nature **375**(6528): 193.

Nussberger, J., G. Wuerzner, *et al.* (2002). "Angiotensin II suppression in humans by the orally active renin inhibitor Aliskiren (SPP100). Comparison with Enalapril." Hypertension. **39**:e1-e8.

Nutt, R. F., S. F. Brady, *et al.* (1988). "Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease." Proc Natl Acad Sci U S A **85**(19): 7129-33.

Oberlin, E., A. Amara, *et al.* (1996). "The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1 [published erratum appears in Nature 1996 Nov 21;384(6606):288]." Nature **382**(6594): 833-5.

Ogasawara, N. (1985). "Markedly unbiased codon usage in Bacillus subtilis." Gene **40**(1): 145-150.

Ogg, G. S., X. Jin, *et al.* (1998). "Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA." Science **279**(5359): 2103-6.

Ohta, Y., T. Masuda, *et al.* (1988). "Isolation of simian immunodeficiency virus from African green monkeys and seroepidemiologic survey of the virus in various non-human primates." Int. J. Cancer **41**(1): 115-122.

Olshevsky, U., E. Helseth, *et al.* (1990). "Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding." J Virol **64**(12): 5701-7.

Oude Essink, B. B., A. T. Das, *et al.* (1996). "HIV-1 reverse transcriptase discriminates against non-self tRNA primers." J Mol Biol **264**(2): 243-54.

Pantaleo, G., J. F. Demarest, *et al.* (1994). "Major expansion of CD8+ T cells with a predominant V beta usage during the primary immune response to HIV." Nature 370(6489): 463-7.

Pantaleo, G., C. Graziosi, *et al.* (1993). "HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease." Nature **362**(6418): 355-8.

Parada, C. A. and R. G. Roeder (1996). "Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain." Nature **384**(6607): 375-8.

Parikh, S. S., C. D. Putnam, *et al.* (2000). "Lessons learned from structural results on uracil-DNA glycosylase." Mutat Res **460**(3-4): 183-199.

Parikh, U., B. A. Larder, *et al.* (2004). HIV Drug Resistance Database, Nucleic Acids Research.

Park, I.W. and J. Sodroski (1995). "Functional analysis of the vpx, vpr, and nef genes of simian immunodeficiency virus." J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **8**(4):335-44.

Patick, A. K., H. Mo, *et al.* (1996). "Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease." Antimicrob Agents Chemother **40**(2): 292-7.

Pazhanisamy, S., C. M. Stuver, *et al.* (1996). "Kinetic characterization of human immunodeficiency virus type-1 protease-resistant variants." J Biol Chem **271**(30): 17979-85.

Pearl, L. and T. Blundell (1984). "The active site of aspartic proteinases." FEBS Lett **174**(1): 96-101.

Pearl, L. H. (1985). Aspartic proteinases and their inhibitors. V. Kostka. Berlin, De Gruyter**:** 189-195.

Pearl, L. H. (1987). "The catalytic mechanism of aspartic proteinases." FEBS Lett **214**(1): 8-12.

Pearl, L. H. (2000). "Structure and function in the uracil-DNA glycosylase superfamily." Mutat Res **460**(3-4): 165-181.

Pearl, L. H. and W. R. Taylor (1987). "A structural model for the retroviral proteases." Nature **329**(6137): 351-4.

Peng, C., B. K. Ho, *et al.* (1989). "Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity." J Virol **63**(6): 2550-6.

Perelson, A. S., P. Essunger, *et al.* (1997). "Decay characteristics of HIV-1-infected compartments during combination therapy." Nature **387**(6629): 188-91.

Perelson, A. S., A. U. Neumann, *et al.* (1996). "HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time." Science **271**(5255): 1582-6.

Phillips, A. N., C. A. Sabin, *et al.* (1995). "HIV results in the frame. Antiviral therapy." Nature **375**(6528): 195; author reply 198.

Phillips, R. E., S. Rowland-Jones, *et al.* (1991). "Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition." Nature **354**(6353): 453-9.

Phillips, S.E. and B.P. Schoenborn (1981). "Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin." Nature **292**(5818): 81-2.

Phoenix, D. A., A. Pewsey, *et al.* (1995). "Distribution and clustering of rare codons in *Escherichia coli* genes." Biochem Soc Trans **23**(4): 503S.

Piana, S. and P. Carloni (2000). "Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme." Proteins **39**(1): 26-36.

Piana S., P. Carloni, *et al.*, (2002). "Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease." J. Mol Biol. **319**(2):567-83.

Piatak, M., Jr., M. S. Saag, *et al.* (1993). "High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR." Science **259**(5102): 1749-54.

Plattner, J. J. and D. W. Norbeck (1989). "Obstacles to drug development from peptide leads." Drug Discovery Technologies. C. R. Clark and W. R. Moos. Chichester England, Ellis Harwood Ltd. Halstead Press: 92.

Polard, P. and M. Chandler (1995). "Bacterial transposases and retroviral integrases." Mol Microbiol **15**(1):13-23.

Polgar, L. (1987). "The mechanism of action of aspartic proteases involves 'push-pull' catalysis." FEBS Lett **219**(1): 1-4.

Polgar, L. (1988). "The different mechanisms of protease action have a basic feature in common: proton transfer from the attacking nucleophile to the substrate leaving group." Acta. Biochim. Biophys. Hung. **23**(3-4): 207-13.

Poon, D. T., J. Wu, *et al.* (1996). "Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity." J Virol **70**(10): 6607-16.

Poorman, R. A., A. G. Tomasselli, *et al.* (1991). "A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base." J Biol Chem **266**(22): 14554-61.

Post, L.E., G.D. Strycharz, *et al.* (1979) "Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in *Escherichia coli*." Proc. Natl. Acad. Sci. USA **76**(4): 1697-701

Powell, D. J., D. Bur, *et al.* (1996). "Expression, characterisation and mutagenesis of the aspartic proteinase from equine infectious anaemia virus [published erratum appears in Eur J Biochem (1997);**246**(1):258]." Eur J Biochem **241**(2): 664-74.

Power, M. D., P. A. Marx, *et al.* (1986). "Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus." Science **231**(4745): 1567-72.

Preston, B. D., B. J. Poiesz, *et al.* (1988). "Fidelity of HIV-1 reverse transcriptase." Science **242**(4882): 1168-71.

Prodromou, C. and L. H. Pearl (1992). "Recursive PCR: a novel technique for total gene synthesis." Protein Eng **5**(8): 827-9.

Profy, A. T., P. A. Salinas, *et al.* (1990). "Epitopes recognized by the neutralizing antibodies of an HIV-1-infected individual." J Immunol **144**(12): 4641-7.

Quere, L., T. Wenger, *et al.* (1996). "Triterpenes as potential dimerization inhibitors of HIV-1 protease." Biochem Biophys Res Commun **227**(2): 484-8.

Rajagopalan, T. G., W. H. Stein, *et al.* (1966). "The inactivation of pepsin by diazoacetylnorleucine methyl ester (DAN)." J. Biol. Chem. **241**(18): 4295-7.

Raju, B., M.S. Deshpande (1991). "Investigating the stereochemistry of binding to HIV-1 Protease with inhibitors containing isomers of 4-amino-3-hydroxy-5-phenylpentanoic acid." Biochem. Biophys. Res. Commun. **180**:187-190.

Raleigh E.A., R. Trimarchi, *et al.* (1989). "Genetic and physical mapping of the mcrA (rglA) and mcrB (rglB) loci of Escherichia coli K-12." Genetics. **122**(2):279-96.

Rangwala, S. H., R. F. Finn, *et al.* (1992). "High-level production of active HIV-1 protease in *Escherichia coli*." Gene **122**(2): 263-9.

Redfield, R. R. and D. S. Burke (1988). "HIV infection: the clinical picture." Sci Am **259**(4): 90-8.

Reedijk, M., C.A.B. Boucher, *et al.* (1995). "Safety, pharmacokinetics, and antiviral activity of A77003, a C-2 symmetry-based Human Immunodeficiency Virus protease inhibitor." Antimicrob. Agents Chemother. **39**: 1559-1564.

Reich, S. H., M. Melnick, *et al.* (1995). "Protein structure-based design of potent orally bioavailable, nonpeptide inhibitors of human immunodeficiency virus protease." Proc Natl Acad Sci U S A **92**(8): 3298-302.

Reich, S. H., M. Melnick, *et al.* (1996). "Structure-based design and synthesis of substituted 2-butanols as nonpeptidic inhibitors of HIV protease: secondary amide series." J Med Chem **39**(14): 2781-94.

Ren, J., R. Esnouf, *et al.* (1995). "High resolution structures of HIV-1 RT from four RT-inhibitor complexes." Nat Struct Biol **2**(4): 293-302.

Ren, J., R. Esnouf, *et al.* (1995). "The structure of HIV-1 reverse transcriptase complexed with 9-chloro- TIBO: lessons for inhibitor design." Structure **3**(9): 915-26.

Ren, S. and E. J. Lien (1998). "Development of HIV protease inhibitors: a survey." Prog Drug Res **51**: 1-31.

Ren, S. and E. J. Lien (2001). "Development of HIV protease inhibitors: a survey." Prog Drug Res **Spec**: 1-34.

Rice, P., R. Craigie, *et al.* (1996). "Retroviral integrases and their cousins." Curr Opin Struct Biol **6**(1): 76-83.

Rich, D. H. (1985). "Pepstatin-derived inhibitors of aspartic proteinases. A close look at an apparent transition-state analogue inhibitor." J Med Chem **28**(3): 263-73.

Rich, D.H. (1990). "Peptidase inhibitors." In Comprehensive Medicinal Chemistry, C. Hansch, P.G. Sammes, and J.B. Taylor, Eds.; Pergamon Press: Oxford. 391-441.

Richards, A. D., L. H. Phylip, *et al.* (1990). "Sensitive, soluble chromogenic substrates for HIV-1 proteinase." J Biol Chem **265**(14): 7733-6.

Richards, A. D., R. Roberts, *et al.* (1989). "Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases." FEBS Lett **247**(1): 113-7.

Richman, D. D., M. A. Fischl, *et al.* (1987). "The toxicity of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial." N Engl J Med **317**(4): 192-7.

Ringhofer, S., J. Kallen, *et al.* (1999). "X-ray structure and conformational dynamics of the HIV-1 protease in complex with the inhibitor SDZ283-910: agreement of time-resolved spectroscopy and molecular dynamics simulations." J Mol Biol **286**(4): 1147-59.

Rittenhouse, J., M.C. Turon, *et al.* (1990). "Affinity purification of HIV-1 and HIV-2 proteases from recombinant *E.coli* strains using pepstatin-agarose." Biochem Biophys Res Commun **171**(1):60-6.

Ro, S., S. G. Baek, *et al.* (1998). "NMR and topochemical studies of peptidomimetic HIV-I protease inhibitors containing a cis-epoxide amide isostere." Bioorg Med Chem Lett **8**(18): 2423-6.

Roberts, J. D., K. Bebenek, *et al.* (1988). "The accuracy of reverse transcriptase from HIV-1." Science **242**(4882): 1171-3.

Roberts, N.A., J.A. Martin *et al.* (1990). "Rational design of peptide-based HIV Proteinase inhibitors." Science **248**: 358-364.

Robertus, J. D., J. Kraut, *et al.* (1972). "Subtilisin; a stereochemical mechanism involving transition state stabilisattion." Biochemistry **11**(23): 4293-303.

Robinson, M., R. Lilley, *et al.* (1984). "Codon usage can affect efficiency of translation of genes in *Escherichia coli.*" Nucleic Acids Res. **12**(17): 6663-71.

Rodgers, D. W., S. J. Gamblin, *et al.* (1995). "The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1." Proc Natl Acad Sci U S A **92**(4): 1222-6.

Roitt, I., Delves, P.J., (2001) "Riott's Essential Immunology." Blackwell Publishing

Rose, J. R., R. Salto, *et al.* (1993). "Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions." J Biol Chem **268**(16): 11939-45.

Rose, R. B., C. S. Craik, *et al.* (1996). "Three-dimensional structures of HIV-1 and SIV protease product complexes." Biochemistry **35**(39): 12933-44.

Rose, R. B., J. R. Rose, *et al.* (1993). "Structure of the protease from simian immunodeficiency virus: complex with an irreversible nonpeptide inhibitor." Biochemistry **32**(46): 12498-507.

Rosenberg, A. H., E. Goldman, *et al.* (1993). "Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system." J Bacteriol **175**(3): 716-22.

Rosin, C. D., R. K. Belew, *et al.* (1999). "Coevolutionary analysis of resistance-evading peptidomimetic inhibitors of HIV-1 protease." Proc Natl Acad Sci USA **96**(4): 1369-74.

Rozzelle, J. E., D. S. Dauber, *et al.* (2000). "Macromolecular inhibitors of HIV-1 protease. Characterization of designed heterodimers." J Biol Chem **275**(10): 7080-6.

Rutenber, E., E. B. Fauman, *et al.* (1993). "Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design." J Biol Chem **268**(21): 15343-6.

Ryu, S. E., P. D. Kwong, *et al.* (1990). "Crystal structure of an HIV-binding recombinant fragment of human CD4." Nature **348**(6300): 419-26.

Ryu, S. E., P. D. Kwong, *et al.* (1990). "Crystal structure of an HIV-binding recombinant fragment of human CD4." Nature **348**(6300): 419-26.

Sakurai, M. M., H. Sugano, *et al.* (1993). "Studies of HIV-1 Protease inhibitors. 1. Incorporation of a reduced peptide, simple amino-alcohol, and statine analogue at the scissile site of substrate sequences." Chem. Pharm. Bull. **41**:1369-1377.

Salto, R., L.M. Babé, *et al.* (1994). "*In vitro* characterisation of nonpeptide irreversible inhibitors of HIV proteases." J. Biol. Chem. **269**(14): 10691-98.

Sattentau, Q. J., J. P. Moore, *et al.* (1993). "Conformational changes induced in the envelope glycoproteins of human and simian immunodeficiency virus by soluble receptor binding." J. Virol. **64:** 7383–7383.

Scharer, O. D. and J. Jiricny (2001). "Recent progress in the biology, chemistry and structural biology of DNA glycosylases." BioEssays **23**: 270-281.

Schechter, I. and A. Berger (1967). "On the size of the active site in proteases." Biochem. Biophys. Res. Com. **27**: 157-162.

Scheiner (1992). In: "Proton Transfer in Hydrogen Bonded Systems." Ed. T. Bontis, Plenum Press, NY.

Schinazi, R. F., B. A. Larder, *et al.* (1997). "Mutations in retroviral genes associated with drug resistance." Int. Antiviral News **5**: 129-142.

Schmalzbauer, E., B. Strack, *et al.* (1996). "Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding *in vitro*." J Virol **70**(2): 771-7.

Schmitz, J. E., M. J. Kuroda, *et al.* (1999). "Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes." Science **283**(5403): 857-60.

Schneider, J. and S. B. Kent (1988). "Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease." Cell **54**(3): 363-8.

Schnittman, S. M., H. C. Lane, *et al.* (1988). "Characterization of GP120 binding to $CD4^+$ and an assay that measures ability of sera to inhibit this binding." J Immunol **141**(12): 4181-6.

Schnolzer, M., H. R. Rackwitz, *et al.* (1996). "Comparative properties of feline immunodeficiency virus (FIV) and human immunodeficiency virus type 1 (HIV-1) proteinases prepared by total chemical synthesis." Virology **224**(1): 268-275.

Schock, H. B., V. M. Garsky, and L.C. Kuo. (1996). "Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity." J Biol Chem **271**(50): 31957-63.

Schoenborn BP (1969). "Neutron diffraction analysis of myoglobin" Nature **224**(215): 143-6.

Schrag, S. J., V. Perrot, *et al.* (1997). "Adaptation to the fitness costs of antibiotic resistance in Escherichia coli." Proc R Soc Lond B Biol Sci **264**(1386): 1287-91.

Schramm, H. J., A. Billich, *et al.* (1993). "The inhibition of HIV-1 protease by interface peptides." Biochem Biophys Res Commun 194(2): 595-600.

Schramm, H. J., J. Boetzel, *et al.* (1996). "The inhibition of human immunodeficiency virus proteases by interference peptides." Antiviral Res. **30**: 155-170.

Schramm, H. J., G. Breipohl, *et al.* (1992). "Inhibition of HIV-1 protease by short peptides derived from the terminal segments of the protease." Biochem Biophys Res Commun **184**(2): 980-5.

Schramm, H. J., H. Nakashima, *et al.* (1991). "HIV-1 reproduction is inhibited by peptides derived from the N- and C-termini of HIV-1 protease." Biochem Biophys Res Commun **179**(2): 847-51.

Schramm, H. J., L. Quere, *et al.* (1998). "Steroids as possible inhibitors of HIV-1 protease." Aids **12**(6): 682-5.

Schubert, U., S. Bour, *et al.* (1996). "The two biological activities of Human Immunodeficiency Virus Type 1 Vpu protein involve two separable structural domains." J. Virol. **70**(2):809-819.

Schubert, U. and K. Strebel (1994). "Differential activities of the Human Immunodeficiency Virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments." J.Virol. **68**:2260-2271.

Seelmeier, S., H. Schmidt, *et al.* (1988). "Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A." Proc Natl Acad Sci U S A **85**(18): 6612-6.

Sellers, T. (1982). "CDC warns of possible pathogen as AIDS cause." Emerg Dep News **4**(12): 11.

Sham, H. L., D. A. Betebenner, *et al.* (1991). "Potent HIV-1 protease inhibitors with antiviral activities in vitro." Biochem Biophys Res Commun **175**(3): 914-9.

Sham, H. L., D. A. Betebenner, *et al.* (1993). "Pseudo-symmetrical difluoroketones. Highly potent and specific inhibitors of HIV-1 protease." FEBS Lett **329**(1-2): 144-6.

Shimotohno, K. and H. M. Temin (1981). "Evolution of retroviruses from cellular movable genetic elements." Cold Spring Harb Symp Quant Biol **45**(Pt 2): 719-30.

Shu, W., J. Liu, *et al.* (2000). "Helical interactions in the HIV-1 gp41 core reveal structural basis for the inhibitory activity of gp41 peptides." Biochemistry **39**(7): 1634-42.

Silva, A. M., R. E. Cachau, *et al.* (1996). "Inhibition and catalytic mechanism of HIV-1 aspartic protease." J Mol Biol **255**(2): 321-46.

Silva, A. M., R.E. Cachau, *et al.* (1995). "Molecular dynamics of HIV-1 protease in complex with a difluoroketone-containing inhibitor: implications for the catalytic mechanism." Adv. Exp. Med. Biol. **362**: 451-454.

Slee, D. H., J. H. Laslo, *et al.* (1995). "Selectivity in inhibition of HIV and FIV protease: inhibitory and mechanistic studies of pyrrolidine-containing a-keto amide and hydroxyethylamine core structures." J. Am. Chem. Soc. **117**: 11867-11878.

Smith, H. and J. R. Kimmel (1960). "Papain (with a section on ficin)." The enzymes. P. D. Boyer, H. Lardy and K. Myrback. New York, Academic Press. **4:** 133-173.

Smolarsky, M. (1978). "Mechanism of action of papain: aryldehydroalanines as spectroscopic probes of acyl enzyme formation." Biochemistry **17**(22): 4606-15.

Smolarsky, M. (1980). "Mechanism of papain catalysis: studies of active-site acylation and deacylation by the stopped-flow technique." Biochemistry **19**(3): 478-84.

Sorensen, M.A., C.G. Kurland, and S. Pedersen. (1989). "Codon usage determines translation rate in *E.coli.*" J. Mol. Biol. **207**: 365-77.

Sorensen, M.A. and S. Pedersen (1991). "Absolute *in vivo* translation rates of individual codons in *Escherichia coli.* The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate." J. Mol. Biol. **222**(2): 265-80.

Speck, R. F., K. Wehrly, *et al.* (1997). "Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop." J Virol **71**(9): 7136-9.

Spina, C. A., H. E. Prince, *et al.* (1997). "Preferential replication of HIV-1 in the CD45RO memory cell subset of primary CD4 lymphocytes in vitro." J Clin Invest **99**(7): 1774-85.

Srinivasakumar, N., M. L. Hammarskjold, *et al.* (1995). "Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that

affect particle formation and Gag- Pol precursor incorporation." J Virol **69**(10): 6106-14.

Stadtman, E. R. and C. N. Oliver (1991). "Metal-catalyzed oxidation of proteins. Physiological consequences." J Biol Chem **266**(4): 2005-8.

Stanton, A. (2003) (1). "Therapeutic potential of renin inhibitors in the management of cardiovascular disorders." Am. J. Cardiovasc. Drugs. **3**(6):389-94.

Stanton, A. (2003) (2). "Potential of renin inhibition in cardiovascular disease." J. Renin Angiotensin Aldosterone Syst. **4**(1):6-10.

Stanton, A., C. Jensen, *et al.* (2003). "Blood pressure lowering in essential hypertension with an oral renin inhibitor, aliskiren." Hypertension. **42**(6):1137-43.

Starcich, B. R., B. H. Hahn, *et al.* (1986). "Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS." Cell **45**(5): 637-48.

Stebbins, J. and C. Debouck (1994). "Expression systems for retroviral proteases." Methods Enzymol **241**: 3-16.

Storer, A. C. and R. Menard (1994). "Catalytic mechanism in papain family of cysteine peptidases." Methods Enzymol **244**: 486-500.

Strebel, K., T. Klimkait, *et al.* (1988). "A novel gene of HIV-1, vpu, and its 16-kilodalton product." Science **241**:1221-3.

Strebel, K., T. Klimkait, *et al.* (1989). "Molecular and biochemical analysis of Humna Immunodeficiency Type 1 Vpu protein." J. Virol. **63**:3784-91.

Strickler, J.E., J. Gorniak, *et al.* (1989). "Characterisation and autoprocessing of precursor and mature forms of human immunodeficiency virus type 1 (HIV-1) protease purified from *Escherichia coli*." Proteins **6**(2): 139-54.

Subramanian, E., M. Liu, *et al.* (1977). "The crystal structure of an acid protease from Rhizopus chinensis at 2.5 A resolution." Adv. Exp. Med. Biol. **95**: 33-41.

Suguna, K., E. A. Padlan, *et al.* (1987). "Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action." Proc Natl Acad Sci U S A **84**(20): 7009-13.

Summers, D. F. and J. V. Maizel, Jr. (1968). "Evidence for large precursor proteins in poliovirus synthesis." Proc Natl Acad Sci U S A **59**(3): 966-71.

Summers, M. F., L. E. Henderson, *et al.* (1992). "Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1." Protein Sci **1**(5): 563-74.

Swain, A. L., M. M. Miller, *et al.* (1990). "X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor." Proc Natl Acad Sci U S A **87**(22): 8805-9.

Sweet, R.W., Truneh, A. and Hendrickson,W. A (1991). "CD4: its structure, role in immune function and AIDS pathogenesis, and potential as a pharmacologicla target." Curr. Opin. Biotech. **2:** 622–633.

Szelke, M., B. Leckie, *et al.* (1982). "Potent new inhibitors of human renin." Nature **299**(5883): 555-7.

Takahashi, M., T. T. Wang, *et al.* (1974). "Acyl intermediates in pepsin and penicillopepsin catalysed reactions." Biochem. Biophys. Res. Commun. **57**(1): 39-46.

Tam, T.F., J. Carriere, *et al.* (1992). "Intriguiing structure activity relations underlie the potent inhibition of HIV protease by norstatine based peptides." J. Med. Chem. **35**: 1318-1320.

Tan, K., J. Liu, *et al.* (1997). "Atomic structure of a thermostable subdomain of HIV-1 gp41." Proc Natl Acad Sci U S A **94**(23): 12303-8.

Tang, J. (1971). "Specific and irreversible inactivation of pepsin by substrate-like epoxides." J. Biol. Chem. **246**(14): 4510-7.

Taylor, M. A. J., K. A. Pratt, *et al.* (1992). "Active papain renatured and processed from insoluble recombinant propapain expressed in *Escherichia coli*." Protein Eng. **5**: 455-459.

Taylor, W. R. (1997). "Residual colours: a proposal for aminochromography." Protein Engineering **10**(7): 743-746.

Terwilliger, E.F., E.A. Cohen, *et al.* (1989). "Functional role of Humna Immunodeficiency Virus type 1 Vpuprotein." J.Virol.**63**:3784-91

Thali, M., J. P. Moore, *et al.* (1993). "Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding." J Virol **67**(7): 3978-88.

Thanki, N., J. K. Rao, *et al.* (1992). "Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: comparisons with molecular modeling." Protein Sci **1**(8): 1061-72.

Thompson, S. K., K. H. Murthy, *et al.* (1994). "Rational design, synthesis, and crystallographic analysis of a hydroxyethylene-based HIV-1 protease inhibitor containing a heterocyclic P1'--P2' amide bond isostere." J Med Chem **37**(19): 3100-7.

Thompson, W. J., P. M. Fitzgerald, *et al.* (1992). "Synthesis and antiviral activity of a series of HIV-1 protease inhibitors with functionality tethered to the P1 or P1' phenyl substituents: X-ray crystal structure assisted design." J Med Chem **35**(10): 1685-701.

Tobin J.B., S.A. Whitt, C.S. Cassidy, P.A. Frey (1995). "Low-barrier hydrogen bonding in molecular complexes analogous to histidine and aspartate in the catalytic triad of serine proteases." Biochemistry. **34**(21):6919-24.

Todd, M. J., N. Semo, *et al.* (1998). "The structural stability of the HIV-1 protease." J Mol Biol **283**(2): 475-88.

Toh, H., R. Kikuno, *et al.* (1985). "Close structural resemblance between putative polymerase of a Drosophila transposable genetic element 17.6 and pol gene product of Moloney murine leukaemia virus." Embo J **4**(5): 1267-72.

Tomasselli, A. G., C. A. Bannow, *et al.* (1992). "Chemical synthesis of a biotinylated derivative of the simian immunodeficiency virus protease. Purification by avidin affinity chromatography and autocatalytic activation." J Biol Chem **267**(15): 10232-7.

Tomasselli, A. G., M. K. Olsen, *et al.* (1990). "Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease." Biochemistry **29**(1): 264-9.

Tong H, and L. Davis (1995). "2-Amino-3-ketobutyrate-CoA ligase from beef liver mitochondria: an NMR spectroscopic study of low-barrier hydrogen bonds of a pyridoxal 5'-phosphate-dependent enzyme." Biochemistry. **34**(10):3362-7.

Tong, L., S. Pav, *et al.* (1993). "Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures." Proc Natl Acad Sci U S A **90**(18): 8387-91.

Tozser, J., F. H. Yin, *et al.* (1997). "Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit." Eur J Biochem **244**(1): 235-41.

Tristem M, C. Marshall, *et al.*, (1992). "Evolution of the primate lentiviruses: evidence from vpx and vpr." EMBO J. **11**(9):3405-12.

Umezawa, H., T. Aoyagi, *et al.* (1970). "Pepstatin, a new pepsin inhibitor produced by Actinomycetes." J Antibiot (Tokyo) **23**(5): 259-62.

Vacca, J. P., B. D. Dorsey, *et al.* (1994). "L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor." Proc Natl Acad Sci USA **91**(9): 4096-100.

Valler, M. J., J. Kay, *et al.* (1985). "The interaction of aspartic proteinases with naturally-occurring inhibitors from actinomycetes and Ascaris lumbricoides." J Enzyme Inhib **1**(1): 77-82.

Veerapandian, B., J. B. Cooper, *et al.* (1992). "Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases." Protein Sci **1**(3): 322-8.

Vernet, T., D. C. Tessier, *et al.* (1990). "Secretion of functional papain precursor from insect cells. Requirement for N-glycosylation of the pro-region." J Biol Chem **265**(27): 16661-6.

Vernet, T., H. E. Khouri, *et al.* (1991). "Processing of the papain precursor. Purification of the zymogen and characterization of its mechanism of processing." J Biol Chem **266**(32): 21451-7.

Wada, K., Y. Wada, *et al.* (1992). "Codon usage tabulated from the genebank genetic sequence data." Nucleic Acids Res **20**: 2111-2118.

Wain-Hobson, S. (1993). "The fastest genome evolution ever described: HIV variation in situ." Curr Opin Genet Dev **3**(6): 878-83.

Wain-Hobson, S. (1993). "Viral burden in AIDS." Nature **366**(6450): 22.

Wain-Hobson, S. (1995). "AIDS. Virological mayhem." Nature **373**(6510): 102.

Wain-Hobson, S., P. Sonigo, *et al.* (1985). "Nucleotide sequence of the AIDS virus, LAV." Cell **40**(1): 9-17.

Wallqvist, A., G. W. Smythers, *et al.* (1998). "A cooperative folding unit in HIV-1 protease. Implications for protein stability and occurrence of drug-induced mutations." Protein Engineering **11**(11): 999-1005.

Wan, M. and B. N. Loh (1995). "Expression and purification of active form of HIV-1 protease from *E.coli*." Biochem. Mol. Biol. Int. **35**: 899-912.

Wang, J. H., Y. W. Yan, *et al.* (1990). "Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains." Nature **348**(6300): 411-8.

Wang, T. T. and T. Hofmann (1976). "Effects of secondary binding by activator and inhibitor peptides on covalent intermediates of pig pepsin." Biochem. J. **153**(3): 701-12.

Wang, Y. X., D. I. Freedberg, *et al.* (1996). "Mapping hydration water molecules in the HIV-1 protease/DMP323 complex in solution by NMR spectroscopy." Biochemistry **35**(39): 12694-704.

Wang, Y. X., D. I. Freedberg, *et al.* (1996). "Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272." Biochemistry **35**(31): 9945-50.

Warshel A., A. Papazyan, P.A. Kollman (1995). "On low-barrier hydrogen bonds and enzyme catalysis." Science. **269**(5220):102-6.

Warshel A., A. Papazyan (1996). "Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds." Proc. Natl. Acad. Sci. U S A. **93**(24):13665-70.

Weber, I. T. (1990). "Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases." J Biol Chem **265**(18): 10492-6.

Weber, I. T., M. Miller, *et al.* (1989). "Molecular modeling of the HIV-1 protease and its substrate binding site." Science **243**(4893): 928-31.

Weber, J. and S. Galpin (1995). "HIV results in the frame. Cyclosporin A." Nature **375**(6528): 198; author reply 198.

Wei, X., S. K. Ghosh, *et al.* (1995). "Viral dynamics in human immunodeficiency virus type 1 infection." Nature **373**(6510): 117-22.

Weiss, A., H. Hollander, and J. Stobo (1985). "Acquired immunodeficiency syndrome: epidemiology, virology, and immunology." Annu Rev Med **36**: 545-62.

Weiss, R. A. (1993). "How does HIV cause AIDS?" Science **260**(5112): 1273-9.

Weissenhorn, W., A. Dessen, *et al.* (1997). "Atomic structure of the ectodomain from HIV-1 gp41 [see comments]." Nature **387**(6631): 426-30.

Wertman, K.F., A.R. Wyman, D. Botstein, *et al.* (1986). "Host/vector interactions which affect the viability of recombinant phage lambda clones." Gene. **49**(2):253-62.

West, M. L. and D. P. Fairlie (1995). "Targeting HIV-1 protease: a test of drug-design methodologies." Trends Pharmacol Sci **16**(2): 67-75.

Wild, C., J. W. Dubay, *et al.* (1994). "Propensity for a leucine zipper-like domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus- induced fusion rather than assembly of the glycoprotein complex." Proc Natl Acad Sci U S A **91**(26): 12676-80.

Wild, C., T. Oas, *et al.* (1992). "A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition." Proc Natl Acad Sci U S A **89**(21): 10537-41.

Wild, C. T., D. C. Shugars, *et al.* (1994). "Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection." Proc Natl Acad Sci U S A **91**(21): 9770-4.

Wilderspin, A. F. and R. J. Sugrue (1993). "Crystallisation and preliminary X-ray investigation of recombinant simian immunodeficiency virus proteinase." J. Mol. Biol. **231**: 1139-1142.

Wilderspin, A. F. and R. J. Sugrue (1994). "Alternative native flap conformation revealed by 2.3 A resolution structure of SIV proteinase." J Mol Biol **239**(1): 97-103.

Wiley, R.A. and D.H. Rich (1993). "Peptidomimetics derived from natural products. Med. Res. Rev. **13**: 327-384.

Willenbrock, F. and K. Brocklehurst (1985). "A general framework of cysteine-proteinase mechanism deduced from studies on enzymes with structurally different analogous catalytic-site residues Asp-158 and -161 (papain and actinidin), Gly-196 (cathepsin B) and Asn-165 (cathepsin H). Kinetic studies up to pH 8 of the hydrolysis of N-alpha-benzyloxycarbonyl-L-arginyl-L-arginine 2-naphthylamide catalysed by cathepsin B and of L-arginine 2-naphthylamide catalysed by cathepsin H." Biochem J **227**(2): 521-8.

Willetts, K. E., F. Rey, *et al.* (1999). "DNA repair enzyme Uracil DNA Glycosylase is specifically incorporated into Human Immunodeficiency Virus Type 1 Particles through a Vpr-Independant Mechanism." J. Virol. **73**(2): 1682-1688.

Willey, R. L., J. S. Bonifacino, *et al.* (1988). "Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160." Proc Natl Acad Sci U S A **85**(24): 9580-4.

Willey, R.L., F. Maldarelli, *et al.* (1992). "Humna Immunodeficiency Virus type 1 Vpu protein induces rapid degradation of CD4." J.Virol. **66**:7193-7200.

Williams, K. J. and L. A. Loeb (1992). "Retroviral reverse transcriptases: error frequencies and mutagenesis." Curr Top Microbiol Immunol **176**: 165-80.

Witte, O. N. and D. Baltimore (1978). "Relationship of retrovirus polyprotein cleavages to virion maturation studied with temperature-sensitive murine leukemia virus mutants." J Virol **26**(3): 750-61.

Wlodawer, A. and J. W. Erickson (1993). "Structure-based inhibitors of HIV-1 protease." Annu Rev Biochem **62**: 543-85.

Wlodawer, A., A. Gustchina, *et al.* (1995). "Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus." Nat Struct Biol **2**(6): 480-8.

Wlodawer, A., M. Miller, *et al.* (1989). "Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease." Science **245**(4918): 616-21.

Wlodawer, A., J. Walter, *et al.* (1984). "Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II." J Mol Biol **180**(2): 301-29.

Wolfenden, R. (1999). "Conformational aspects of inhibitor design: enzyme-substrate interactions in the transition state." Bioorg Med Chem **7**(5): 647-52.

Wolinsky, S. M., B. T. Korber, *et al.* (1996). "Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection." Science **272**(5261): 537-42.

Wollert, K.C., and H. Drexler, (1999). "The renin-angiotensin system and experimental heart failure." Cardiovascular Res. **43**:838-849.

Wondrak, E.M., J.M. Louis, *et al.* (1991). "Purification of HIV-1 wild-type protease and characterisation of proteolytically inactive HIV-1 protease mutants by pepstatin A affinity chromatography." FEBS Lett **280**(2): 347-50.

Wong, J. K., M. Hezareh, *et al.* (1997). "Recovery of replication-competent HIV despite prolonged suppression of plasma viremia." Science **278**(5341): 1291-5.

Wong-Staal, F., B. H. Hahn, *et al.* (1984). "Molecular characterization of human T-lymphotropic leukemia virus type III associated with the acquired immunodeficiency syndrome." Princess Takamatsu Symp **15**: 291-300.

Wood, J.M., F. Cumin, *et al.* (1994). "Pharmacology of renin inhibitors and their application to the treatment of hypertension." Pharmacol. Ther. **61**(3):325-44.

Wood, J.M., J. Maibaum, *et al.* (2003). "Structure-based design of aliskiren, a novel orally effective renin inhibitor." Biochem. Biophys. Res. Comm. **308**:698-705.

Woon, T. C., R. I. Brinkworth, *et al.* (1992). "Inhibition of HIV-1 proteinase by metal ions." Int J Biochem **24**(6): 911-4.

Wright, S. (1932). "The roles of mutation, inbreeding, crossbreeding, and selection in evolution." Proceedings of the Sixth International Congress on Genetics.

Wu, L., N. P. Gerard, *et al.* (1996). "CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5." Nature **384** (6605): 179-83.

Wu, H., Kwong, P. D. and Hendrickson, W. A. (1997). "Dimeric association and segmental variability in the structure of human CD4". Nature **387:** 527–530.

Wurtz, A. and E. Bouchat (1879). "Sur le ferment digestif du papaya." Compt. Rend. **89**: 425-430.

Wyatt, R., P. D. Kwong, *et al.* (1998). "The antigenic structure of the HIV gp120 envelope glycoprotein." Nature **393**(6686): 705-11.

Yamazaki, T., A. P. Hinck, *et al.* (1996). "Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy." Protein Sci **5**(3): 495-506.

Yamazaki, T., L. K. Nicholson, *et al.* (1994). "Secondary structure and signal assignments of human-immundodeficiency-virus-1 protease complexed to a novel, structure-based inhibitor." Eur. J. Biochem. **219**(1-2): 707-12.

Yang, X., C. H. Herrmann, *et al.* (1996). "The human immunodeficiency virus Tat proteins specifically associate with TAK *in vivo* and require the carboxyl-terminal domain of RNA polymerase II for function." J Virol **70**(7): 4576-84.

Yaron, A., A. Hatzubai, *et al.* (1998). "Identification of the receptor component of the IkappaBalpha-ubiquitin ligase." Nature **396** (6711): 590-4.

Yoshinaka, Y. and R. B. Luftig (1980). "Physicochemical characterisation and specificity of the murine leukemia virus Pr65gag proteolytic factor." J. Gen. Virol. **48**(Pt2): 329-40.

Yu, Z., P. Caldera, *et al.* (1996). "Irreversible Inhibition of the HIV-1 Protease: Targeting Alkylating Agents to the Catalytic Aspartate Groups." J. Am. Chem. Soc. **118**: 5846-5856.

Zennou, V., F. Mammano, *et al.* (1998). "Loss of viral fitness associated with multiple Gag and Gag-Pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo." J Virol **72**(4): 3300-6.

Zhang, C. T. and K. C. Chou (1994). "An alternate-subsite-coupled model for predicting HIV protease cleavage sites in proteins." Protein Eng **7**(1): 65-73.

Zhang, Y. M., H. Imamichi, *et al.* (1997). "Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites." J Virol **71**(9): 6662-70.

Zhang, Z.-Y., R. A. Poorman, *et al.* (1991). "Dissociative inhibition of dimeric enzymes. Kinetic characterisation of the inhibition of HIV-1 protease by its COOH-terminal tertrapeptide." J. Biol. Chem. **266**: 15591-15594.

Zhang, Z. Y., I. M. Reardon, *et al.* (1991). "Zinc inhibition of renin and the protease from human immunodeficiency virus type 1." Biochemistry **30**(36): 8717-21.

Zhao, B., E. Winborne, *et al.* (1993). "Three-dimensional structure of a simian immunodeficiency virus protease/inhibitor complex. Implications for the design of human immunodeficiency virus type 1 and 2 protease inhibitors." Biochemistry **32**(48): 13054-60.

Zhao Q., C. Abeygunawardana, *et al.* (1996). "NMR evidence for the participation of a low-barrier hydrogen bond in the mechanism of delta 5-3-ketosteroid isomerase." Proc Natl Acad Sci U S A. **93**(16):8220-4.

Zhao Q, C. Abeygunawardana, *et al.* (1997). "Hydrogen bonding at the active site of delta 5-3-ketosteroid isomerase." Biochemistry. **36**(48):14616-26.

Zhou, Q. and P. A. Sharp (1996). "Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV- 1 Tat." Science **274**(5287): 605-10.

Zutshi, R., J. Franciskovich, *et al.* (1997). "Targeting the dimerisation interface of HIV-1 protease: inhibition with cross-linked interfacial peptides." J. Am. Chem. Soc. **119**: 4841-4845.